How important are model structural and contextual uncertainties when estimating the optimized performance of water resource systems?

Barnaby Dobson[1], Thorsten Wagener[1,2], Francesca Pianosi[1]

[1]Departartment of Civil Engineering, University of Bristol, Bristol, UK

[2]Cabot Institute, University of Bristol, UK

Corresponding author: Barnaby Dobson (barnaby.dobson1@gmail.com)

Key Points

- We measure the impact of epistemic structural and contextual uncertainty on the estimated performance of optimized operating policies.
- We show contextual uncertainty (around definition of the system boundaries) has the most significant impact on the estimated performances.
- Performance of optimized management strategies can only be estimated properly if aleatory uncertainty is also adequately accounted for.

Index terms

Reservoirs, decision making under uncertainty, system operation and management, multi-objective optimization, water management

Abstract

Uncertainty in simulating water resource systems (WRSs) makes it difficult to assess how effective different water management decisions will be. Uncertainty in simulation models can undermine the credibility of simulation and optimization studies and the uptake of their results. We identify different sources of uncertainty in WRS models and find that structural uncertainty (i.e. around definition of interrelationships within the system) and contextual uncertainty (i.e. around definition of the system boundaries) are rarely considered when simulating and optimizing WRSs. We propose a methodology to quantify the effects of structural and contextual uncertainties on the estimated performance of optimized water management decisions and demonstrate that they have a significant impact on a real-world case study of a pumped-storage system in the UK. To the best of the authors' knowledge this is the first study to consider the impact of these types of uncertainty on optimized operating policies and their simulated performances. Our main finding is that, of all the considered uncertainties, the assumptions made about context – specifically around the level of cooperation between neighbouring water companies – had the greatest impact on performance estimates. This is important because few WRSs exist in isolation, yet discussion of the effects that a given definition of the system boundaries have on the simulation/optimization results is uncommon. We also highlight the significance of adequately considering aleatory uncertainty when evaluating performance estimates – something that few studies do – and present a simple technique to justify the sample size used for the evaluation of optimization results.

1 Introduction

Models and model-based optimization are widely used to support decision making in water resource management. Within the broad area of water resource management, this study will focus on the optimization of reservoir operations (Yakowitz, 1982; Yeh, 1985; Hiew et al., 1989; Labadie, 2004; Rani and Moreira, 2009). Reservoir operation optimization typically refers to identifying the operational decisions (or the operating policies to make those decisions based on the system conditions, such as reservoir storage or time of year) that achieve optimal values of certain objectives (for example, reliability of water supply) . Objectives are evaluated using a numerical model that simulates the interaction between decisions and forcing inputs (for example, demands and reservoir inflows) over time. Optimization can be particularly beneficial in systems of interconnected reservoirs, where even a relatively small increase in system complexity can make the definition of effective operating policies far from trivial (Moss et al., 2016). Another difficulty in making operational decisions is the need for balancing multiple conflicting objectives, which 30% of large dams face worldwide (ICoLD, 2003). When multiple objectives exist, the aim of optimization is not to find a single optimal operating policy, but rather to characterise the possible trade-offs within a set of candidate policies (Cohon and Marks, 1975; Haimes and Hall, 1977; Guariso et al., 1986; Kasprzyk et al., 2013; Reed and Kollat, 2013).

Although reservoir operation optimization methods have been extensively studied in the scientific literature, their application in practice is limited and many authors over time have pointed out a disconnect between research and practice in this field (Yeh, 1985; Simonovic, 1992; Labadie, 2004; Brown et al., 2015). The first attempt to survey the uptake of reservoir optimization tools (Rogers and Fiering, 1986) revealed that the uncertainty present in the underpinning simulation models contributed to a significant lack of trust in the end results of the optimization process. In a more recent survey on the perception of uncertainty by water managers, Höllermann and Evers (2017) found that the uncertainty around boundary conditions, which is an example of what we will later define as contextual uncertainty, was the source of uncertainty of highest concern for practitioners. In the context of a climate change impacts study, Mahmoud et al. (2009) found that stakeholders did not trust the study results if they were not convinced by the system conceptualization underlying the simulation models used in the assessment. In order to increase the credibility and hence the use of optimization results in practice, we believe that it is essential to understand to what extent optimized solutions will maintain their estimated level of performance in the face of the uncertainties that unavoidably affect the simulation model used during optimization.

A common conceptual classification of uncertainties affecting simulation models distinguishes between *aleatory* and *epistemic* uncertainty (Walker et al., 2003; Beven et al., 2017). Aleatory uncertainty arises from intrinsic random variability in the system, such as variability in weather conditions. It is typically considered irreducible, but it can be well characterised statistically. Epistemic uncertainty instead can be defined as the uncertainty that is attributable to a lack of historical observations (Beven et al., 2017), which results in a lack of understanding about the system, its properties and its expected behaviour (Walker et al., 2003) and is difficult to characterise in statistical terms. Examples are the uncertainty in the projected magnitude of a flood event with return period exceeding the length of historical time series, or the uncertainty in the subsurface properties of a catchment, which are typically not observable. Epistemic uncertainty is in principle reducible, even if this may be difficult to do in practice. Below we discuss how aleatory and epistemic uncertainties affect WRS simulation models, and we review the techniques that have been used to address them within optimization studies.

Practically unavoidable in WRSs is the variability in hydrological forcing, such as inflows into reservoirs, which was the main focus of the earliest water management studies (Maass et al., 1962). A common practice in the field is to assume that inflows are aleatory and stationary processes, although the validity of the stationarity assumption is highly debated (Milly et al., 2008; Montanari and Koutsoyiannis, 2014), and represent them by fitting a statistical model to the historic data (see for example Matalas (1967) for an early application and Vogel (2017) for a recent review of the available techniques). Reservoir operation is then stochastically optimized under this statistical model, for example via Stochastic Dynamic Programming (e.g. (Stedinger et al., 1984; Nardini et al., 1992; Castelletti et al., 2012)) or by generating a synthetic sequence of forcing for which the operations are deterministically optimized (Koutsoyiannis and Economou, 2003). The more densely sampled the statistical model (either by using a high-resolution discretization grid for Stochastic Dynamic Programming or by generating a long sequence of synthetic forcing), the longer the optimization will take. Hence, it is good practice to keep the sample size limited during the optimization process and then re-evaluate the optimized operations over an expanded sample, so to get a more reliable estimate of their performances (Mortazavi et al., 2012). Similar considerations apply to other system variables that can be regarded as aleatory uncertainties, such as water demand and evaporation from reservoir surfaces, which are often modelled using similar statistical models to inflows (Donkor et al., 2014).

As for epistemic uncertainties, we distinguish four types: parametric, objective, structural and contextual. Parameters are constant values in a model, typically identified through measurement or calibration (Walker et al., 2003). The measurement and calibration processes not being exact, it results in a certain amount of uncertainty in parameter values, which is in principle reducible by further measurement and testing. However, in WRS simulation and particularly for long-term evaluation of WRS performance, it is typically necessary to use *conceptual* parameters that do not relate to specific physical quantities but instead encapsulate and simplify complex and diverse phenomena. Examples are trend parameters that summarise long-term changes in water demand or in inflow statistics as a consequence of climate change. In recent years, several studies have investigated the robustness of WRS management solutions to epistemic uncertainty in parameters, either by including sampling of the uncertain parameter space in the re-evaluation of optimized operating policies (Kasprzyk et al., 2012; Herman et al., 2014) or by directly incorporating the sampling of uncertain parameters into the optimization process (Trindade et al., 2017; Watson and Kasprzyk, 2017).

Another source of epistemic uncertainty is the choice and formulation of *model outcomes* (Walker et al., 2003) such as, in the case of WRS optimization, the metrics of system performance (or *objectives* hereafter). It is well known that similar formulations (for example, vulnerability vs reliability) of the same objective (for example, water supply) may yield different suggested operations (Hashimoto et al., 1982). A further difficulty is that the decision-makers themselves may not be aware of their true preferences until they are able to visualise operating policies and their respective objectives in the context of the trade-offs available to them (Kasprzyk et al., 2013). Unlike parametric uncertainty, which can be characterised by random sampling within a defined range of plausible parameter values, the geometry of the plausible objective space is more challenging to define. Recently, Quinn et al. (2017) presented a method to investigate the effects of competing formulations of uncertain objectives on multi-objective optimization results. This method creates different framings of the WRS management problem using different objective formulations, where each framing is regarded as a single sample in the space of uncertain objectives. By application to a hydropower reservoir system in Vietnam, Quinn et al. (2017) found that the choice of objective has a significant impact on how effective an operating policy would be considered.

Another source of uncertainty, and a key focus of this study, is model structural uncertainty. We use the definition of (Walker et al., 2003), who suggests that model structural uncertainty is uncertainty about *"the behaviour of the system and the interrelationships among its elements"*. Examples in WRS management might include the type of statistical model used to describe aleatory variables or the omission of processes that are poorly understood or unsupported by data, such as pump failures. The effects of structural uncertainty on the prediction of environmental or socio-economic variables has been relatively well studied, for example in hydrological (Clark et al., 2008), water quality (Beck, 1987), ecological (Ayala et al., 2014) and water distribution system (Hutton and Kapelan, 2015) modelling. However, to the best of the authors' knowledge, it has not yet been considered in any detail for its impact on optimized solutions of WRS management problems. Instead, the structural choices underlying simulation models used in this field often seem to be guided by a lack of data or knowledge, or by the need to make a certain optimization method applicable (computationally tractable), rather than their appropriateness (Giuliani et al., 2015b).

Finally, we list a source of uncertainty that is rarely considered in WRS modelling: contextual uncertainty. (Walker et al., 2003) defines it as the uncertainty about *"the boundaries of the system to be modelled"*. Since few WRS exist in isolation, a certain degree of contextual uncertainty is unavoidable, just like in the modelling of any open system (Dooge, 1973). Typical examples of contextual uncertainties in WRS modelling include aggregating demand nodes beyond the chosen system boundary, and assuming cooperation between multiple operators in the same system. We focus on this last element specifically because it is common for optimization studies to assume that if multiple infrastructures are present in the same system their operations are perfectly coordinated, However, in reality there often are different operators that either do not coordinate their decisions or do so through ad hoc discussions rather than formal rules that can be represented within a simulation model (Giuliani et al., 2015a). Central to this point, a growing number of studies have demonstrated that cooperation in water systems is a critically important factor in improving operational decisions (Tilmant and Kinzelbach, 2012; Anghileri et al., 2013; Giuliani and Castelletti, 2013; Marques and Tilmant, 2013; Wu et al., 2016).

Our study makes three contributions. Firstly, we introduce and assess a workflow to measure the impact of model structural and contextual uncertainties on the estimated performance of WRS management solutions obtained by optimization. The workflow enables modellers to assess whether optimization results are robust to uncertainty in their underlying simulation models. It transfers the 'rival framings' framework, presented by (Quinn et al., 2017), to address the relevance of structural and contextual modelling choices in estimating the performance of the solutions of a multi-objective optimization problem. Secondly, our study demonstrates the value of this workflow in a specific case study of a two-reservoir pumped storage system. In this application we answer the question "What is the extent to which the performances of optimized reservoir operating policies change upon re-evaluation under a simulation model that makes different modelling choices?" or, more simply: "How worse/better off can performances be when optimized under a modelling choice that turns out to be incorrect?". As we later discuss, the conclusions we draw from this case study application are, in varying degrees, generalizable to other types of WRS optimization problems. Thirdly, we test the importance of aleatory uncertainty in the re-evaluation of optimized operating policies. Previous studies (e.g. Mortazavi et al. (2012)) suggested that optimal solutions obtained by using insufficient realisations of aleatory uncertainties (e.g. short climate records) can be severely flawed. Recent studies (Kasprzyk et al., 2013; Herman et al., 2014; Quinn et al., 2017) have used expanded sampling strategies to better account for aleatory uncertainty when re-evaluating optimization results. Here, we will present and apply a simple approach to justify the chosen sample size for re-evaluation.

The paper is organised as follows. In Section 2 we introduce our workflow to quantify the impact of structural and contextual uncertainty on estimates of performance of optimized operating policies. In Section 3 we describe our case study model, the formulation and experimental setup for the optimization problem (complemented by the full set of model equations in the Appendix) and introduce our technique to determine the adequate sample size for re-evaluation. In Section 4 we present the results of applying our methodology to the case study model and in Section 5 we discuss the implications and generalizability of these results.

2 Methodology

Modelling choices are often made under significant uncertainty. In a sense, there is a hierarchy among these uncertainties. We draw on Walker [2003] to describe this hierarchy. The choice of the context and of the objectives typically come first, and is determined by a range of factors, from institutional constraints to data availability, and the very question that the decision maker wishes to ask. Within a given model context, different model structures may be selected. Each model structure will have its own set of parameters. Each of these levels may have many sub-levels, for example the choice to use an autoregressive model for synthetic streamflow generation then leads to the choice of the lag in that model. As discussed in the introduction, uncertainty sources may be introduced at any of these levels. We would suggest that, depending on the strength of statistical justification, any of these sources of uncertainty may be regarded as aleatory or epistemic. In the face of so much complexity, we follow the philosophy that once a choice is made for every source of epistemic uncertainty, one hypothesis of the real system is created and the residual aleatory uncertainty can be represented by Monte Carlo simulations.

Our methodology leverages this philosophy to study the impact of, in our case, structural and contextual uncertainties on WRS optimisation results. To this end,-we use a workflow built on the 'Rival Framings' framework introduced by (Quinn et al., 2017). In this workflow, each rival framing is a plausible hypothesis for formulating the WRS management problem. In

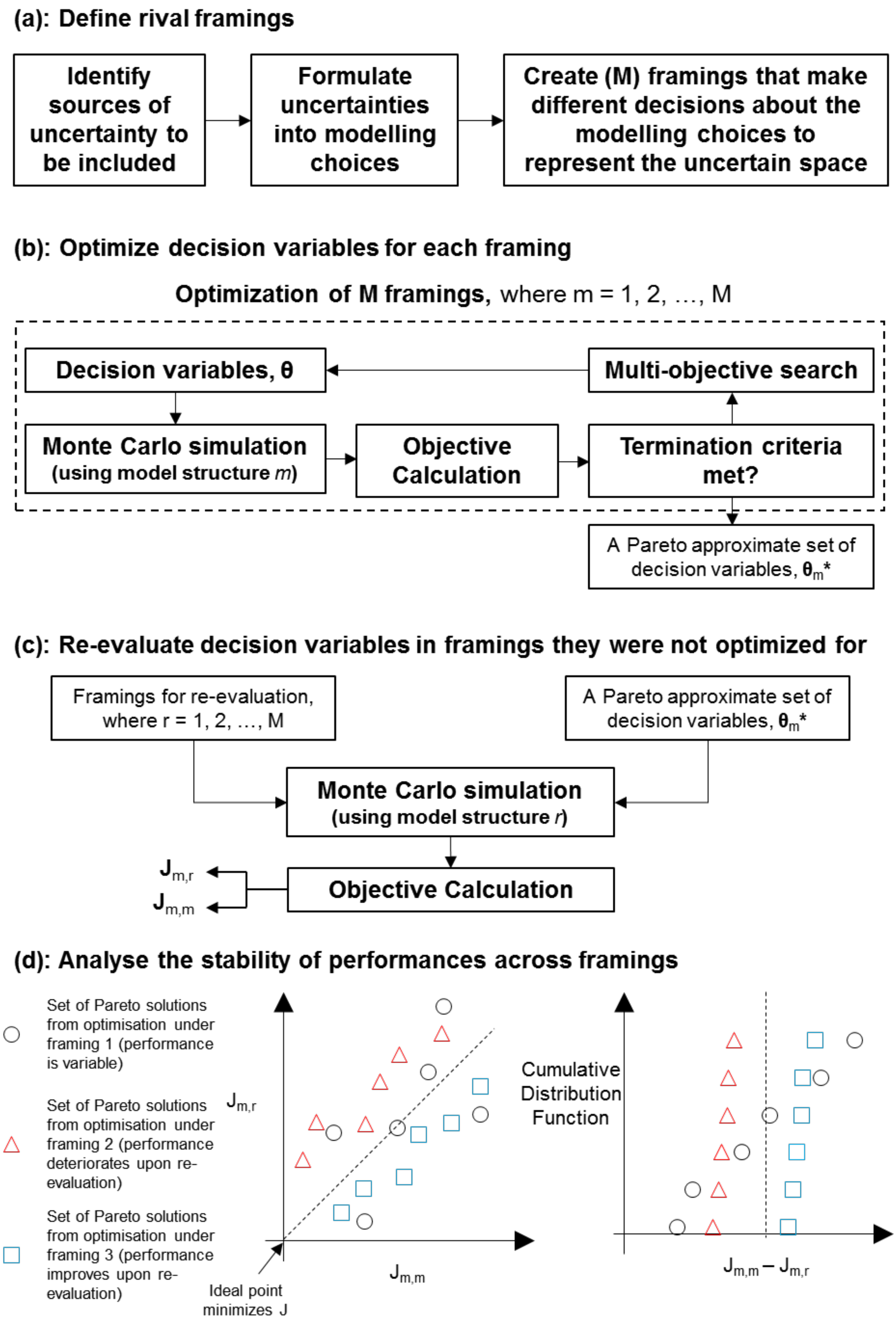
Quinn et al. (2017), each framing uses different formulations of the objectives with no changes in the underlying simulation model. In our study, each framing makes different choices about some elements of the simulation model structure and the context. Figure 1 presents this workflow.

The first step is to define the rival framings, as depicted in Figure 1a. During this step, uncertainties are identified and characterised. For example, in Quinn et al. (2017) the sources of uncertainty are the objectives, which are characterised by a set of different options for their formulation. In our case, the sources of uncertainty are some of the assumptions underlying the model structure (for example, whether to include pump failures) and context (for example, whether the two water companies that manage the two reservoirs in the WRS coordinate their operations). Each framing will then consist of a unique set of modelling choices relating to these uncertain assumptions. Ideally, the range of considered framings should fully represent the uncertain space under investigation. In our case, this means that they should capture the uncertain assumptions in the modelling process that either the decision-maker(s) or the modeller(s) are sceptical about or wish to study their exposure to, in line with the second and sixth principles of best practice in collaborative modelling (Langsdale et al., 2013): *"all stakeholder representatives participate early and often to ensure that all their relevant interests are included"* and "*the model addresses questions that are important to the decision makers and stakeholders"*. This step can also be mapped into the 'identify uncertainties' stage in the XLRM framework presented by (Lempert et al., 2003) and demonstrated in (Lempert et al., 2006; Kasprzyk et al., 2013). The remaining stages of that framework – identify decision 'levers' (L), map actions to outcomes (R) and define performance metrics (M) – should then be followed to formulate a relevant simulation model of the system and thus create an appropriate management problem. As suggested by (Mahmoud et al., 2009), it is important to interact with the decision maker(s) throughout the modelling process since their trust in the model outcomes increases with their trust in the underlying system conceptualization.

Next, as depicted in Figure 1b, decision variables are optimized under each framing, which results in a set of (approximate) Pareto optimal solutions (hereafter, a set of Pareto solutions) for each framing. In our case study, the decision variables are not the operational decisions (reservoir releases and pumped inflows) directly, but rather the parameters defining the operating policies that will be used to compute those decisions based on the WRS state (we discuss this in more detail in the case study section and Appendix 1). A set of Pareto solutions are those whose performance in any objective can only be improved with a corresponding reduction in performance in one or more of the remaining objectives. In order to account for aleatory uncertainties (e.g. in our problem the streamflow, demand and potential evaporation time series) we use Monte Carlo simulation for the calculation of the objective function values against a set of possible realisations of those uncertainties. For the multi-objective search, we use the Borg multi-objective evolutionary algorithm (MOEA) since it has been shown to perform very effectively for multi-objective reservoir operation problems (Salazar et al., 2016). The key working principle of Borg is to blend a range of heuristic optimization approaches, balancing them to adapt as the search progresses. However, any optimization tool capable of robustly solving stochastic, multi-objective formulations could be used here.

The key step in the Rival Framings workflow is the use of an independent re-evaluation of the optimized solutions under different candidate framings (Figure 1c). This is effectively testing how stable the estimated performances and trade-offs are to the assumptions made in the framing used for optimization. As in the optimization step, we use Monte Carlo simulation to estimate the objective values under aleatory uncertainty. Because the aim of this step is to show the stability of estimated performances, it is important that the approximation error from the Monte Carlo simulation be small enough to enable meaningful comparisons between different sets of simulations. Previous studies that performed a re-evaluation step of stochastic optimization results have often used a larger sample size than the one used for optimization, so to reduce approximation error in the re-evaluation (for example, (Kasprzyk et al., 2013; Herman et al., 2014; Giuliani et al., 2015b; Quinn et al., 2017)). Here, we propose linking the re-evaluation sample size with the decision-maker's sensitivity to differences in objective values. For example, if the decision-maker is only sensitive to differences in cost greater than 10 £/day, we should choose a sample size such that the approximation error in the objective calculation is, at a maximum, 10 £/day. Further increasing the sample size would be unnecessary, given that the decision-maker would not discriminate between solutions with cost differences lower than 10 £/day. In the experimental setup section, we will provide a simple technique to implement this idea and derive an adequate sample size for given value of the decision-maker's sensitivity.

Finally, the results of the re-evaluation step can be visualised through an objective-objective plot (shown on the left in Figure 1d), where the performances in the framing used for optimization are plotted against the performances in the framing used for re-evaluation (Quinn et al., 2017). If the points lie along the bisector (the x=y line), there is no difference in performance between the two framings. If instead the points deviate from the bisector, then the choice of the framing impacts performance estimates. The larger the deviations from the bisector, the less robust the solutions are to the modelling choices underpinning the different framings. Given that it is the deviations from the x=y line that are critical to assessing robustness, we propose a simpler visualisation (shown on the right in Figure 1d), which shows the cumulative distribution function (CDF) of these deviations. In this plot, if the performances in the framing used for optimization are similar to the performances in the re-evaluation framing, the CDF would follow the x=0 line. Again, such a result would mean the modelling choices that distinguish the two framings have minimal impact on optimized performances. If instead the CDF lies to the left of the x=0 line (as in the case of the red triangles in Figure 1d), then the framing used for optimization is estimating lower objective values (i.e. more desirable, if we assume all objectives are to be minimized) than in the framing used for re-evaluation – i.e. solutions perform worse upon re-evaluation with respect to what suggested by optimization. If the CDF lies to the right (blue squares) then solutions perform better upon re-evaluation – i.e. the estimates of performance produced by optimization are conservative and likely to be exceeded if the assumptions underpinning the optimization model are wrong. Obviously, the latter situation is preferable than the former, although we would suggest that both outcomes are not satisfying as they reveal a significant amount of uncertainty. Similarly, we would expect decision-makers to be most concerned by sets of Pareto solutions that exhibit the widest variation in performance, such as the black circles in Fig. 1d, since pinning down the expected performance of a given solution under uncertainty will be difficult. Overall, these CDF plots provide decision makers with an indication of how stable individual solutions, and whole sets of Pareto solutions, are likely to be under different sources of uncertainty. This both enables them to select solutions from a set of Pareto solutions with characterised robustness and directs them towards sources of uncertainty whose monitoring and reduction will reduce the variability in performance estimates the most.



(a): Define rival framings
Identify sources of uncertainty to be included
Formulate uncertainties into modelling choices
Create (M) framings that make different decisions about the modelling choices to represent the uncertain space
(b): Optimize decision variables for each framing
Optimization of M framings, where m = 1, 2, …, M
Decision variables, θ
Multi-objective search
Monte Carlo simulation (using model structure m)
Objective Calculation
Termination criteria met?
A Pareto approximate set of decision variables, θm*
(c): Re-evaluate decision variables in framings they were not optimized for
Framings for re-evaluation, where r = 1, 2, …, M
A Pareto approximate set of decision variables, θm*
Monte Carlo simulation (using model structure r)
Objective Calculation
Jm,r
Jm,m
(d): Analyse the stability of performances across framings
Set of Pareto solutions from optimisation under framing 1 (performance is variable)
Set of Pareto solutions from optimisation under framing 2 (performance deteriorates upon re-evaluation)
Set of Pareto solutions from optimisation under framing 3 (performance improves upon re-evaluation)
Jm,r
Jm,m
Ideal point minimizes J
Cumulative Distribution Function
Jm,m – Jm,r

Figure 1. The rival framings workflow used in this study to estimate the impact of model structure and contextual uncertainties on the performance of optimized WRSs management solutions.

3 Case Study and experimental set-up

3.1 Description of case study and simulation model

Figure 2 depicts the water system used in this study to demonstrate our workflow. It is a simplified version of a two-reservoir system in the South West of the UK (labelled as S1 and S2 in Figure 2), with a pumped storage element to provide a supply of water in dry weather (to S1). The system is partly shared between two different water companies, reservoir S1 being the system element used by both companies. This reservoir is used by Company 1 to support downstream abstraction during low river flows for around 400,000 people (demand node D1 in Figure 2). While company 2 uses it to complement releases from S2 in supplying around 150,000 people (D2). The two reservoirs are moderately sized (relative to other large UK reservoirs, that have an average of 1377 Ml, (EA, 2017)) with storage capacities in the order of 20,000 megalitres (Ml) (S1) and 5000 Ml (S2). Both reservoirs must make environmental compensation releases, in addition S1 is occasionally required to deliver larger releases for downstream fisheries. Besides ensuring a reliable supply to D1 and D2, and delivering environmental compensation releases, the reservoirs' operation also aims at minimising pumping costs.

The two companies that operate the system liaise regularly, particularly regarding pumped storage operation, which is constrained by control curves and has operated in eight years since 1995. In simulation exercises for strategic and long-term planning, the two companies do not jointly model the system, instead making agreed conservative assumptions about each other's operation (described in the following section). Decision procedures are negotiated, and individual decisions are made cooperatively in either emergency situations or as periods of dry weather extend, informed by wider system considerations such as regional demand. Hence, for simulation purposes, the definition of system boundaries and the degree of cooperation assumed in the model of this system is a good example of contextual uncertainty.

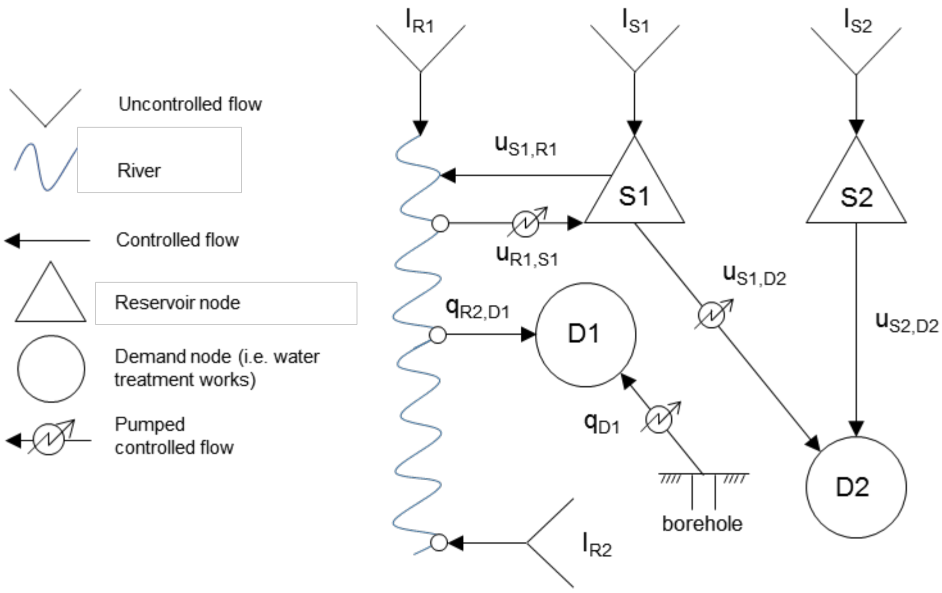


Figure 2. A schematic showing the main components and flows of the two-reservoir system used as a case study. Two companies operate half of the system each, with one reservoir (S1) as a shared resource. Company 1 takes the release and abstraction decisions $u_{R1,S1}$, $u_{S1,R1}$ and q_{D1} with the aim of supplying D1, while Company 2 makes the release decisions $u_{S1,D2}$ and $u_{S2,D2}$ to supply D2. Reservoir inflows are described by I_{S1} and I_{S2}, and river streamflows by I_{R1} and I_{R2}.

In our simulation model, we use ‘operating policies’ to represent the decision-making behaviour of the reservoir operators. Operating policies are parameterised functions that take the system’s state variables as inputs (for example, reservoir storages and inflows) and return operational decisions as outputs (i.e. the three releases denoted by $u_{S1,R1}$, $u_{S1,D2}$ and $u_{S2,D2}$, and the pumped flux $u_{R1,S1}$ in Figure 2). As further explained in the next section, the choice of whether to use one operating policy to produce all operational decisions at once, or a separate policy for each reservoir, depends on the assumed degree of cooperation in the model. In either case, the parameters of the operation policies are the decision variables in our optimization problem. This approach to reservoir operation optimization is common in the water resource systems literature, e.g. (Guariso et al., 1986; Oliveira and Loucks, 1997; Koutsoyiannis and Economou, 2003; Quinn et al., 2017) and is a specific instance of the more general ‘direct policy search’ approach to dynamic systems optimization (Rosenstein and Barto, 2001). For the parameterised functions, it is quite common to use universal approximators because they should be sufficiently flexible to approximate the (unknown) optimal operating policy under any given scenario. For example, artificial neural networks (Pianosi et al., 2011) and radial basis function networks (Giuliani et al., 2015b) have both been applied in this field – we use the latter here. The reservoirs’ mass balance equations and model forcing inputs (i.e. reservoir inflows, evaporation losses and water treatment work demands) are resolved at a daily time-step. A mathematical description of the simulation model, including the operation policies, can be found in Appendix 1. The model forcing inputs are generated synthetically using statistical models trained on historical data. We describe this process in detail in Appendix 2. The model is coded in the C programming language and parallelised with the Open MPI library, which enables highly efficient simulation runs.

3.2 Definition of modelling choices and framings

The model structure and contextual uncertainties considered in our rival framings methodology are summarized in Figure 3. These uncertainties lead to different possible modelling choices for generating inflows, demand and evaporation, and for representing pump failures, fisheries releases and the level of cooperation between the two companies. In this study, we consider each modelling choice as a binary option. For the inflow, demand, evaporation and fisheries releases, the binary choice is between a more sophisticated or less sophisticated representation of the process. For the pump breaks and cooperation between companies, the choice is between including them in the model or not. While we recognise that limiting our study to such binary choices may reduce some of the nuances in our interpretation, we believe it is useful to assess how important the choice is before investigating it in detail: if the exclusion of a process has no impact on the results, it is unlikely that the specific formulation of that process will matter. Then, we combine these binary choices to formulate 8 progressively more complex framings. In the following paragraphs, we briefly describe these sources of uncertainty and associated modelling choices, while further mathematical details are provided in Appendix 2.

Rival framing	Modelling choices					
	Pump breaks occur	Inflow model	Demand model	Potential evaporation model	Fisheries allowable period	Company cooperation
1:NoBrk-AR1-Sep-UC	×	AR(1)	AR(1)	AR(1)	September	×
2:Brk-AR1-Sep-UC	✓	AR(1)	AR(1)	AR(1)	September	×
3:Brk-AR2-Sep-UC	✓	AR(2)	AR(2)	AR(2)	September	×
4:Brk-AR2-All-UC	✓	AR(2)	AR(2)	AR(2)	All year	×
5:NoBrk-AR1-Sep-C	×	AR(1)	AR(1)	AR(1)	September	✓
6:Brk-AR1-Sep-C	✓	AR(1)	AR(1)	AR(1)	September	✓
7:Brk-AR2-Sep-C	✓	AR(2)	AR(2)	AR(2)	September	✓
8:Brk-AR2-All-C	✓	AR(2)	AR(2)	AR(2)	All year	✓

Figure 3. The set of rival framings used in this study. Each row indicates a different framing and the modelling choices associated with it. AR(1)/AR(2) describes the type of autoregressive statistical model used to generate the input (with lag of 1 day or 2 respectively). Fisheries releases may occur in September or year-round (except Spring). Company cooperation indicates that all objectives (and decisions) in the system are optimized together while non-cooperation indicates that each company's objectives (and decisions) are optimized separately.

Pump breaks. Based on communication with operators at the two water companies, there is significant uncertainty associated with pump failures resulting from pump breaks. They occur infrequently, for unique reasons and in unique ways.~~.~~ Based on the authors' interactions with the operators of this and other water systems in the UK, pump failures account for some of the largest operational failures that water suppliers face and are a key cause of the practitioners' scepticism about the value of simulation models, which typically neglect them.

Uncertainty in modelling pump failures exists around both the choice of the statistical distribution of occurrence, severity and duration of failures (structural uncertainty) and the choice of the parameters of those distributions (parametric uncertainty). For the purpose of this study, we limit the choice in our framings to whether to include pump failures or not. While this may simplify the aforementioned uncertainties, it will at least indicate whether including pump failures in the WRS model significantly affects the optimization results or not. When pump breaks are included, we chose to represent both their duration and the duration between consecutive breaks by a Poisson distribution because this is commonly used for characterising failure frequency in systems with electronic components (Weiss, 1956).

Inflow, demand and evaporation. The characterisation of reservoir inflow uncertainty dates to early works in the field (Maass et al., 1962) and has been an active field of research since (Fiering and Bund, 1971; Hirsch, 1979; Salas et al., 2005; Rajagopalan et al., 2010; Herman et al., 2016). Here we assume that inflows are stationary and thus, using the definitions set out in our introduction, they can be fully represented through statistical models (see Vogel (2017) for a discussion of the non-stationary case). This representation introduces a source of aleatory uncertainty but also of parametric uncertainty (associated with the parameters of the statistical model) and structural uncertainty (associated with the choice and form of the statistical model). The latter is particularly significant when the quality and quantity of historic data that can be used to fit the statistical model is low, as in our case. Here we consider two possible model structures to generate inflows: both are periodic autocorrelated (AR) models (Salas and Obeysekera, 1982) but with different lags (of 1 and 2 days). Further details on the AR models and their calibration from historical data are given in Appendix 2. As for demand and evaporation, we note that many techniques that address inflow uncertainty may broadly be applied to these variables too, as demonstrated by the similarity between the methods described in (Donkor et al., 2014) for statistical modelling of water demand and those described in the references above for statistical modelling of inflows. Hence, we also use lag-1 and lag-2 AR models for evaporation and demand (see Appendix 2).

Fisheries releases. Reservoir S1 is occasionally required by the UK Environment Agency to make a large release over a few days to support downstream fisheries. Predicting when such request may occur is difficult because it depends on the decision made by an external stakeholder, the Environment Agency, who acts according to conditions and demands (for example, downstream water quality and pressure from the fisheries' owners) that occur outside the boundaries of the WRS under study. This problem of how to represent forcing inputs driven by the behaviour of external stakeholders is rather common in WRSs modelling and makes the fisheries releases a good example of contextual uncertainties. In their simulation exercises for long-term planning, Company 1 assumes that fisheries releases may only occur over a period in September, since this is the most common time of year for them to occur. However, the historic data shows that the releases have also occurred at many other times of the year, apart from the Spring period. In this study we thus characterise this uncertainty by starting the fisheries releases on a random day inside a feasible time window, which is either limited to September (as in the Company's assumption) or expanded to the entire year except Spring (as in the historic data). Once the starting date has been randomly selected, the overall volume of water released is fixed (the historic data shows that the total volume released each year is relatively constant) and distributed over a period of random duration between 3 and 15 days (the historic data determines these limits).

Representation of company cooperation. The last modelling choice we consider is how to represent cooperation between the two companies when making release or pumping decisions. We include this choice because previous studies (e.g. (Tilmant and Kinzelbach, 2012; Anghileri et al., 2013; Giuliani and Castelletti, 2013; Marques and Tilmant, 2013; Wu et al., 2016)) demonstrated that model assumptions about coordination between connected reservoirs can dramatically impact the performance of optimized reservoir operations. We note that the current situation is that the two companies coordinate their operations but do so on a case-by-case basis accounting for conditions in the wider water resources system of which the two reservoirs are part. For these reasons it is difficult to formalise their current coordination of operations into a set of mathematical equations. On the other hand, in the simulation models that the two companies use for long-term planning, these elements of cooperation are represented by a set of conservative assumptions about the other company's operations, which is a precaution deemed acceptable as the system is typically in surplus. In order to capture the uncertainty around the assumed level of cooperation in modelling this system, we thus define two extreme scenarios: a 'non-cooperative' modelling scenario and a 'cooperative' one. In the 'cooperative' modelling scenario the model simulates the entire two-reservoir system as one WRS (as depicted in Figure 2), and the optimization produces one operation policy that returns all the release and pumping decisions simultaneously. In the 'non-cooperative' modelling scenario, instead, two separate operation policies are produced, each controlling the company's own decisions independently from the other. The two separate simulation models used to optimize the two policies are shown in Figure 4. The model of Company 1, shown on the left, makes the conservative assumption that Company 2 will always draw as much water as possible from reservoir S1, i.e. $u_{S1,D2}$ is equal to the maximum licensed. Conversely, the model of Company 2, on the right, assumes that they will always be able to draw their licensed volume from reservoir S1. We note that whilst these two scenarios capture the two possible extremes of modelling the system, the actual operation varies somewhere between, depending on the nature of the situation.

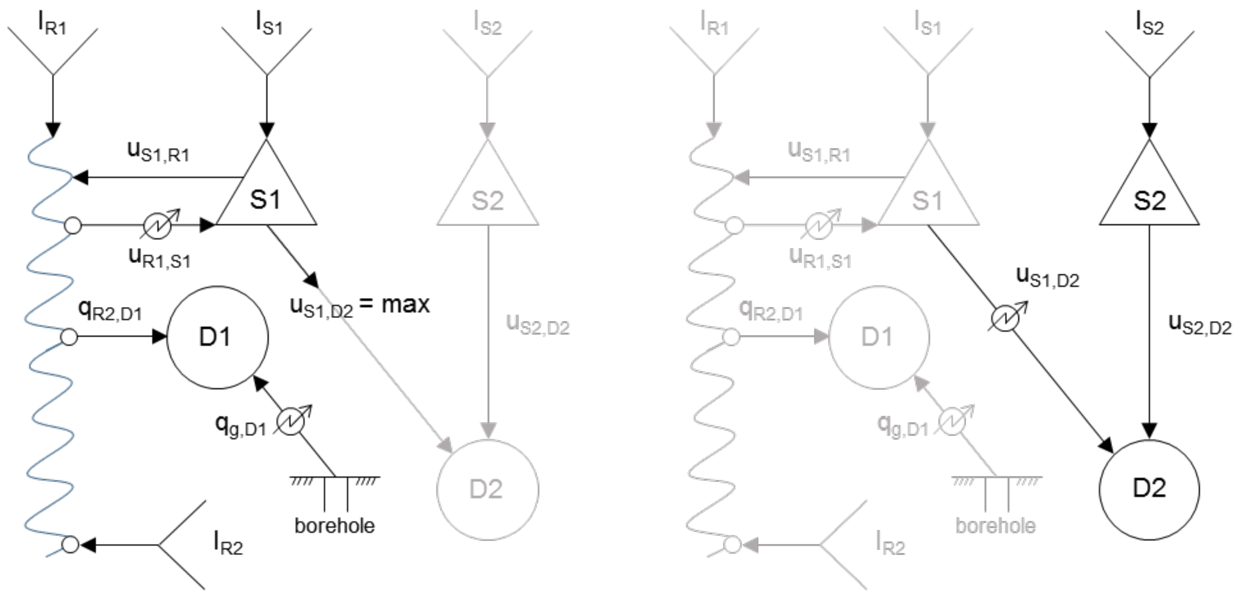


Figure 4. Schematic of the two separate simulation models used in the 'non-cooperative' modelling scenario.

3.3 Optimization and re-evaluation

Both the optimization and re-evaluation steps of our methodology (Figure 1 (b) and (c)) require calculation of the objectives via Monte Carlo simulation. We have already defined the operator's objectives broadly: each company aims to reduce both deficits in supply and the cost of pumping. This gives four objectives: the average daily deficit in supply for company 1 (J_{D1}) and for company 2 (J_{D2}); and the average daily pumping cost for company 1 ($J_{£1}$) and company 2 ($J_{£2}$). The mathematical details of the objectives formulation are given in Appendix 1. Quinn et al. (2017) has demonstrated that the objective formulation is important and that, for example, the optimal operations can be different even just for a change in the temporal statistic used to aggregate costs, e.g. the mean, the worst case or another distribution quantile. However, this is not the focus of our study and so we will not explore the impact of using different objective function formulations. For each objective we will take simulated daily costs and average them across both the simulation period and the Monte Carlo realisations, as shown in Figure 5 and summarised in equation (1):

$$J_i = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{T}\sum_{t=1}^{T} g_{t,i}^{k}\left(\theta, p_t^k, I_t^k\right), \quad (1)$$

where i is the objective index (running from 1 to 4 in our case), k is the index of the Monte Carlo realisations, t is the time index (day); g are the daily costs of operation (i.e. supply deficits at the two demand nodes and daily pumping costs for the two companies), θ are the searched for parameters of the operation policy, p and I are uncorrelated and autocorrelated forcing respectively, T is the length (days) of the simulation period and K is the number of runs in the Monte Carlo simulation. In the optimization step, we aim to find the set of θ vectors that is the set of Pareto solutions between the 4 objectives. We use the Borg MOEA (Hadka and Reed, 2013) to solve the optimization problem since it has been shown to be highly effective for reservoir operation optimization problems based on direct policy search (Salazar et al., 2016).

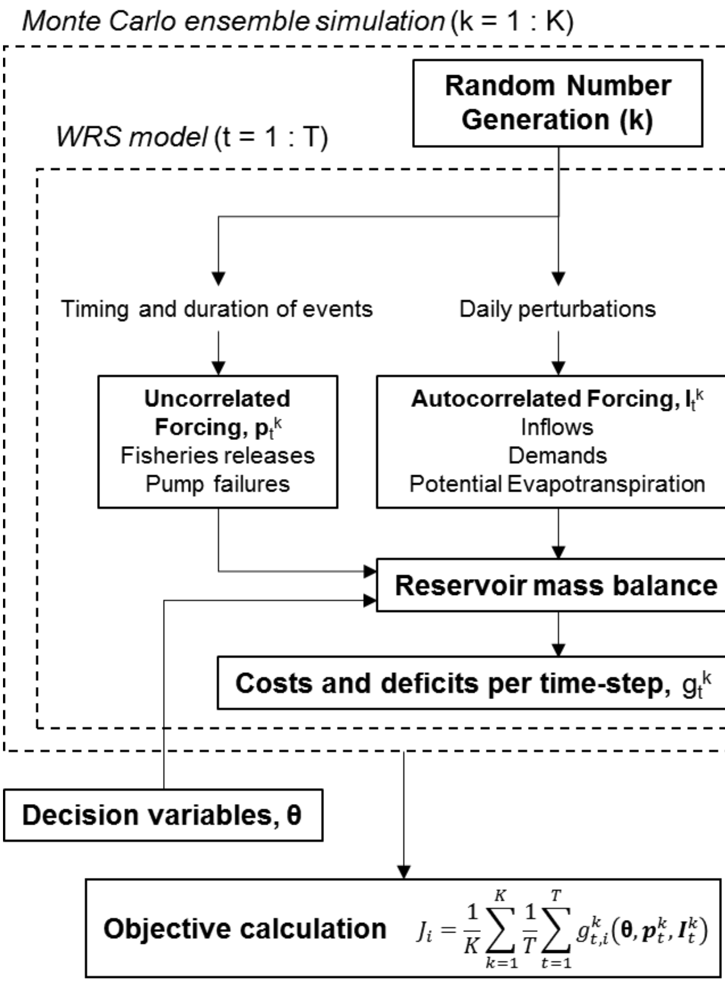


Figure 5. The Monte Carlo simulation process behind evaluating the objectives associated with a given operation policy. k is the index for the Monte Carlo ensemble member, and t is the index for the time-step.

As anticipated above, the Monte Carlo simulation requires specification of the ensemble size K and of the simulation length T – together the ‘sample size’. A given sample size will have an associated approximation error in the objective values. As the sample size tends to infinity, the approximation error should reduce to zero. Thus, for a given operation policy, the approximation error at a given sample size can be quantified as the difference between an objective value at that sample size and the objective value at an infinite sample. Given the high computational efficiency of our simulation code, we approximate the objective value at an infinite sample by the objective value at a very large sample size ($4*10^7$ days, coming from K = 400 and T = 100,000). We can then monitor the trajectory of approximation errors at other sample sizes. An example for the water deficit objective of company 2 is given in Figure 6.

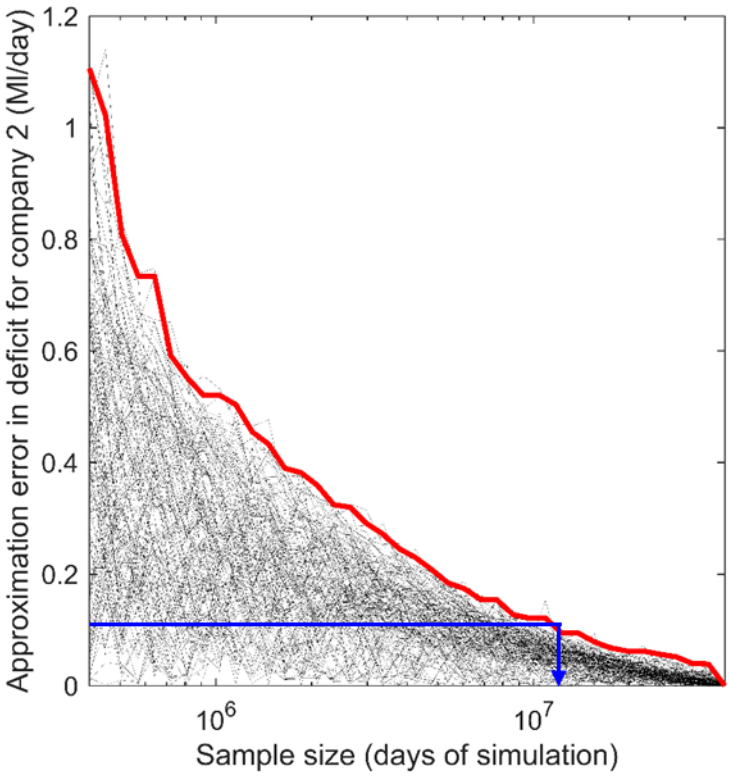


Figure 6. (Black lines) approximation error of 132 random policies evaluated over simulation periods of independent and increasingly large sample sizes. (Red line) the 99th percentile of these errors at each given sample size. (Blue line) an example of how to start with the sensitivity of the decision maker for this objective (0.11Ml/d) and determine an appropriate sample size ($1.2*10^7$ days). Approximation errors are defined as the absolute difference between the objective function value at a given sample size and the value at the largest sample size possible (in this case, $4*10^7$ days).

During the optimization step, smaller sample sizes will result in less accurate objective evaluations (as is clear from Figure 6) but speedier computation, thus enabling more function evaluations during optimization. A vast literature on 'noisy optimization' indeed shows that optimization algorithms can often find good solutions in spite of approximate objective values (Fitzpatrick and Grefenstette, 1988; Miller and Goldberg, 1996; Smalley et al., 2000; Yun et al., 2010). Salazar et al. (2017) provide an extensive discussion and computational experiments to demonstrate the complex trade-off between objective approximation and optimization efficacy. Additionally, using the Borg MOEA requires specification of an 'epsilon' value, which is the minimum difference in objective value that must be exceeded for the optimizer to consider one solution to outperform another. Kasprzyk et al. (2012) show that this epsilon value should reflect the minimized likelihood of one solution being selected over another due to approximation error for the given sample size.

During re-evaluation, we must ensure that simulation results have approximation errors smaller than the decision-maker's 'sensitivity', i.e. their ability to discriminate between different solutions. As exemplified by the blue line in Figure 6, we can start from the pre-specified sensitivity of the decision maker and calculate the sample size that would guarantee approximation errors below that sensitivity. With this approach and assuming sensitivities of 0.17 Ml/day for J_{D1}, 0.11Ml/day for J_{D2}, £9/day for $J_{£1}$ and £1.3/day for $J_{£2}$, we obtain here a required sample size of K = 400, T = 30,000, i.e. $1.2*10^7$ days (the plots equivalent to the 99th percentile in Figure 6 for all objectives and all framings are given in Appendix 3, Figure A3.1). Because our simulation model is computationally highly efficient, for the optimization step we simply use the same sample size as in the re-evaluation step. Consistent with the interpretation of epsilon values given by (Kasprzyk et al., 2012), we set epsilons to the decision maker sensitivities given above.

With epsilon values specified, the last tuning parameter to be specified for running Borg MOEA is the maximum number of function evaluations. To make this choice, we repeat the optimization process multiple times with different seeds to determine an appropriate number of function evaluations to produce a stable hypervolume. A hypervolume indicates the volume of objective space that is captured by a set of Pareto solutions, as described in (Knowles and Corne, 2002). An example plot, for framing 8, can be found in Appendix 3, Figure A3.2. This allows us to conclude that an optimization process with 10^5 iterations should be more than satisfactory.

Finally, for framings that include no cooperation between companies, we need to implement the following small adjustments to the workflow, to account for the specifics of our case study:

- *In the optimization step*. Each company has its own model and its own operation policy that is completely independent from the other company's model and policy. This results in two separate optimizations for a single framing, one for each company (i.e. separate optimizations of the two models depicted in Figure 4). Hence there is no trade-off between Company 2's objectives and Company 1's, given that every operational solution for Company 1 is compatible with every operational solution for Company 2 (and vice versa). We note that, while this may not be true in reality, it is the result of the assumptions made under the non-cooperation modelling choice.
- *In the re-evaluation step*. The policies developed under the cooperative modelling choice use as inputs the state variables from the entire system (i.e. storage at both reservoirs, demands, and all uncontrolled flows). Under the non-cooperation choice, instead, the system is de-coupled, thus it would be impossible to simulate a policy developed in the cooperative case under a non-cooperative assumption since certain inputs to that policy are simply not represented in the two de-coupled models. For example, a policy that depends on the storage in reservoir S2 could not be simulated by the model in the left panel of Figure 4. Consequently, in the re-evaluation step we can only re-evaluate policies in the cooperative framings. Note that non-cooperative policies can instead be simulated in the cooperative framings since they control

different release variables and both policies' input data are represented in the coupled, cooperative model.

4 Results

As anticipated in the Methodology section, our key result is a comparison of the performances of a set of Pareto solutions as estimated in the optimization step and in a re-evaluation step where different model framings are used. We show the results of this comparison for each of the four objectives, and for each combination of Pareto solutions and framings, in Figure A3.3 of Appendix 3. The majority of the subplots in this Figure show large deviations from the x=0 line, which means that the different modelling choices made in the 8 framings have a large impact on estimated objective values. For simplicity of illustration, we discuss in detail the results for a specific objective (the mean deficit in the water supply for Company 2, J_{D2}) and re-evaluation framing (8), shown in Figure 7.

Figure 7 shows the cumulative distribution function (CDF) of the differences between the J_{D2} estimates under the framings used for optimization (from 1 to 8) and under framing 8. The variability in estimated performance for the non-cooperative framings (1, 2, 3, 4 - red lines) is noticeably larger than for the cooperative ones (5 ,6, 7, 8 - blue lines) – indicated by the larger spread over the x-axis. We also see a clear difference between the framings that use an AR(1) model for generating synthetic forcing during optimization (framings 1, 2, 5, 6 - lines with circles) and the ones that use an AR(2) model (3, 4, 7, 8 – no circles). In fact, the CDFs of the AR(1) framings are more commonly positive, meaning that the objective values typically improve when re-evaluated using an AR(2) model, i.e. that using the AR(1) synthetic generator provides a conservative estimate of performances. The differences attributable to other modelling choices (i.e. pump failures and fisheries releases) are far smaller. While the details of these results are specific to objective J_{D2} and framing 8, the conclusions of which modelling choices make a significant difference are similar across all cases, as we show in the expanded results plots in Appendix 3, Figure A3.3.

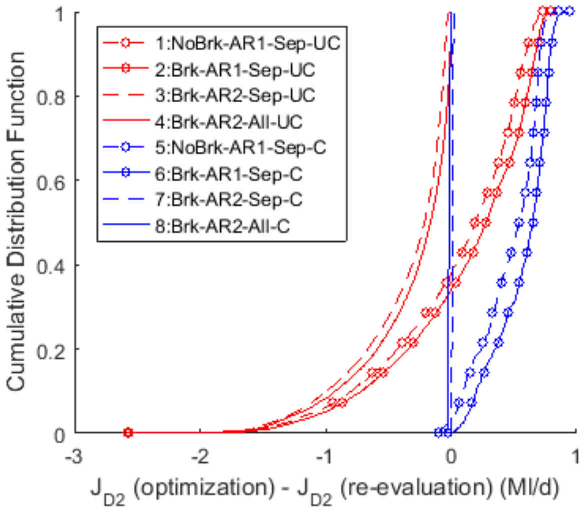


Figure 7. A cumulative distribution function of the differences between the J_{D2} objective values estimated in the optimization step (under framings 1 to 8) and the J_{D2} values re-evaluated under framing 8. Blue lines indicate non-cooperative framings and red lines indicate cooperative framings. Results are obtained with a sample size of $1.2*10^7$ days for both optimization and re-evaluation.

To offer a more detailed interpretation of the impact of contextual uncertainty, we further analyse two of the eight sets of Pareto solutions shown in Figure 7: framing 4 (i.e. a non-cooperative framing) and 8 (i.e. a cooperative framing). From each of these two sets, we extract the subset of solutions that form the set of Pareto solutions between the objectives deficit in water supply for company 2 (J_{D2}) and pumping costs for company 1 ($J_{£1}$) . The objective values of these subsets are shown in Figure 8 as red points (framing 4) and black points (framing 8). From this figure we see that there is an area in which the red points are higher than the black points, i.e. an area where solutions optimized under framing 8 perform systematically better than those optimized under framing 4 in terms of water deficit (J_{D2}), while being equal in terms of pumping costs ($J_{£1}$ between 300 and 1000 £/d). Thus, it is clear that optimizing under framing 4 simply does not allow access to part of the objective trade-off space that could instead be accessed if optimizing under framing 8. This effect is due to the non-cooperative assumption made in framing 4, given that it is persistent across all non-cooperative framings: if a set of policies optimized in any non-cooperative framing is evaluated under any cooperative framing, they will always lose access to this section of the objective subspace, which is instead accessible to all sets of policies optimized under any cooperative framing.

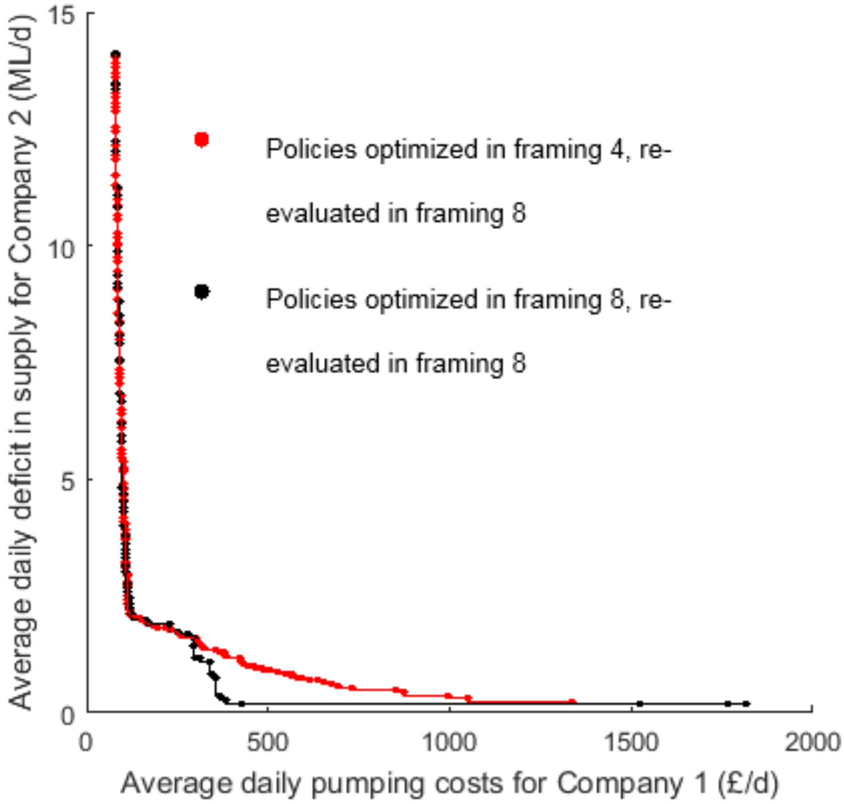


Figure 8. (Red points) estimated performances of the set of Pareto solutions optimized under framing 4 and re-evaluated using framing 8. Only policies that lie on the trade-off between the pump costs for company 1 (X-axis) and deficit for company 2 (Y-axis) are shown for clarity. (Black points) the same but with the Pareto set of operation policies optimized in framing 8.

The results in Figure 7 were obtained by using a very large sample size ($1.2*10^7$ days) in both optimization and re-evaluation. Using such a large sample size to re-evaluate optimized solutions is not common as the majority of studies in this field use simulation periods of few years or decades – with some exceptions such as (Kasprzyk et al., 2013; Herman et al., 2014; Quinn et al., 2017). Hence, we thought it helpful to repeat our analysis using a smaller sample size of 10^4 days, which corresponds to a more typical simulation length of 30 years. We present a sample of the results so obtained in Figure 9, again for objective J_{D2} and re-evaluation in framing 8. From this Figure, it seems that all sources of uncertainty are influential. However, a closer inspection of the results shows that even the estimates of performance for framing 8 vary from optimization to re-evaluation (i.e. the CDF deviates from the x = 0 line). Since the only difference from optimization and re-evaluation here is the Monte Carlo sample used for simulation, we can attribute the differences in estimated performance to approximation errors from the small sample size. Therefore, we expect that Figure 9 shows the combined influence of both modelling choices and approximation errors. Inspecting these results in their entirety (Figure A3.4 in Appendix 3), shows that there is no discernible pattern to the influence of different uncertainties, signifying that the role of aleatory uncertainty is possibly as large as that of modelling choices.

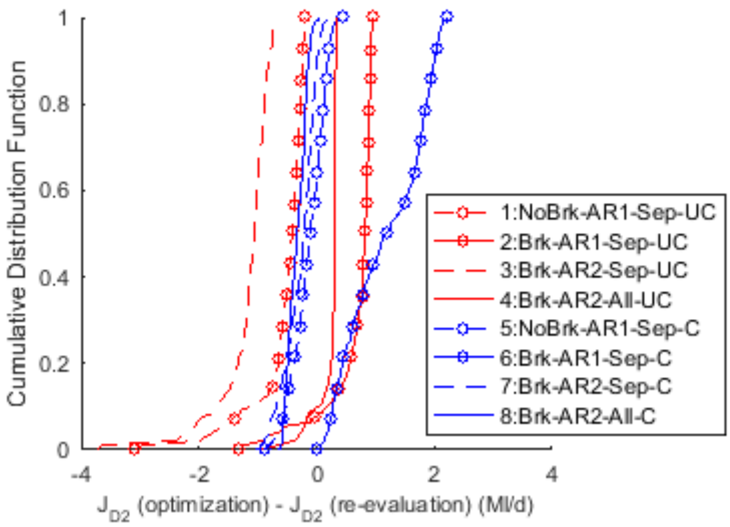


Figure 9. The same as Figure 7 but with an optimization and re-evaluation sample size of $1.1*10^4$ days (i.e. 30 years).

5 Discussion

The key aim of this study was to measure the impact of model structural and contextual uncertainties on the estimated performance of water management decisions obtained by optimization. Our key result is that, in our case study, the assumption about the level of cooperation between water companies has a greater impact on estimating objective values than any other modelling choice. Our estimates of performance vary largely with this assumption, as shown in Figure 7, and if we model either company separately from the other, the benefit of optimization is hindered by the inaccessible trade-off space, as shown in Figure 8. This confirms what other studies have found, i.e. that cooperation in decision-making is both impactful and beneficial (Tilmant and Kinzelbach, 2012; Anghileri et al., 2013; Giuliani and Castelletti, 2013; Marques and Tilmant, 2013; Wu et al., 2016). It is important to note that this occurs even though the assumptions made in the non-cooperative scenario about each company's operations are compatible and seemingly conservative. In company 2's simulation model, it is assumed that they can take as much water as they want from the other company's reservoir (S1), while in company 1's model it is assumed that company 2 will take as much as they can from S1. These assumptions expose company 2 to an over-reliance on S1 and results in company 1 over-abstracting from the river in their pumped storage operations, caused by their conservative assumptions about company 2. In turn, this two-reservoir system that we have modelled is part of a larger inter-connected water resource system. Therefore, it is possible that even the most robust results we present here are themselves subject to a similar amount of uncertainty if one considered the larger system. In general, few water systems exist in isolation and thus these contextual uncertainties likely impact the results of many optimization studies. Beven and Alcock (2012) have suggested that the choice of system boundaries is crucial in making predictions about environmental systems, here we have shown that the same is also true for a coupled human-environmental system.

Another conclusion from our results is that the objective values are quite sensitive to the choice of the autoregressive synthetic generators (Figure 7). In Appendix 3, Figure A3.3, the complete set of CDF plots show that the performance estimates of policies optimized under framings 5 and 6 (AR1, cooperative) nearly always improve or stay the same when re-evaluating in framings 7 and 8 (AR2, cooperative). This implies that the AR(1) generator leads to more conservative estimates. That including just one additional autocorrelation lag term impacts the operational performance corroborates the conclusions found in (Tejada-Guibert et al., 1995): that small changes in the statistical characterisation of input forcing can have large operational impacts.

An encouraging result is the seeming lack of importance of the choice of including pump failures, a factor that is often mentioned by practitioners as critically missing in simulation models. For example, in Figure 7 we see only a small translational shift in the CDF between framings 1 (no breaks, non-cooperative) and 2 (breaks, non-cooperative), and between framings 5 (no breaks, cooperative) and 6 (breaks, cooperative). This indicates that, while pump breaks do reduce performance, the choice of the optimal policy is unlikely to change.

This is because, in the event of a pump failure, there is little that can be done in terms of the operational decisions available in the model to tackle the failure. We expect this result to be generalisable since it would require a level of redundancy not usually present in water resources systems to alter the conclusion. The authors hope that more studies will include this rarely considered source of uncertainty on the basis that it may help to build a case that excluding asset failures in a simulation model is not a reason to reject optimized operational policies.

In Figure 9, we show that approximation errors can lead to falsely attributing differences in objective values to (in our case) structural/contextual uncertainty. Despite this, the use of an expanded sample for re-evaluation of optimization results is the exception and not the rule in this field (examples of using expanded re-evaluation samples are (Kasprzyk et al., 2013; Herman et al., 2014; Giuliani et al., 2015b; Quinn et al., 2017)) and studies where a simulation period of 20-30 years (or even less) is used to demonstrate the efficacy of a new optimization algorithm are not unusual. For our case study, Figure 6 shows that this simulation length is far too short to produce accurate objective estimates, and Figure 9 (and the expanded results shown in Appendix 3, Figure A3.4) show that, with a 30-year re-evaluation period, there are seemingly many significant differences between framings, which could be misattributed to structural/contextual uncertainty if we were not aware of how much approximation error was present in the objective value estimates.

6 Conclusion

In this paper we formally investigated the impacts of model structural uncertainty and of contextual uncertainty on optimization results through application to a two-reservoirs system. Our results revealed four key findings, three about the impact of uncertainty and one around the importance of re-evaluation:

1. Cooperation between operators is often assumed in water resources models. We find that this assumption and thus the definition of the system boundaries (i.e. the model's context) had the most significant impact on estimated objective values and trade-offs.
2. The model structural uncertainty that had the most significant impact was around the level of temporal persistence (auto-correlation) in the forcing inputs. Our results suggest that even minor differences in the statistical formulation of forcing generators can significantly impact performance estimates. One implication of this result is that the common simplification of using an AR(1) model for generating forcing inputs (as is often done to reduce the problem's dimensionality for stochastic dynamic programming) may not always be a suitable assumption.
3. Other modelling choices, such as whether to introduce pump failures or not, had much less impact. This result is encouraging since it shows that simplifying assumptions in simulation models do not always affect optimization results significantly and hence simulation-optimization models can be operationally useful even if they are not a perfect picture of the real-world.
4. Recreating results by re-evaluation on a shorter (30 year) simulation period produced dramatically different conclusions – this shows that insufficiently accounting for aleatory uncertainty (i.e. intrinsic random variability) can lead to misleading results.

Our findings highlight why it is important to consider structural and contextual uncertainties in water resources system optimization. Re-evaluation under uncertainty enables decision-makers test how 'optimal' their optimization results really are, and thus identify the modelling choices that merit careful consideration. It can also identify simplifying assumptions that, although seemingly unrealistic, can be acceptable for the purpose of operation optimization.

Acknowledgements

B. Dobson is funded as part of the WISE CDT under a grant from the Engineering and Physical Sciences Research Council (EPSRC), grant number EP/L016214/1. F. Pianosi is partially funded by an EPSRC Living with Environmental Uncertainty Fellowship, grant number EP/R007330/1. The authors are also very grateful to C. Hutton and Wessex Water for the data, time and effort they devoted to this work. Partial support for T. Wagener was provided by a Royal Society Wolfson Research Merit Award. Some of the ideas in this paper benefited from discussions with Professor Patrick Reed and Dr Julianne Quinn during a research visit at Cornell University. The C code to implement the optimization and re-evaluation experiments presented in this paper is available at doi: 10.5281/zenodo.2556998. The data used to calibrate and evaluate the water resource model and the synthetic generators is property of the water companies and as such only modified data has been shared by the authors.

Appendix 1: Simulation and operation of the water resource system

Simulation

The model dynamics are simulated at a daily time-step through the following two mass balance equations

(1) $S_{S2,t+1} = S_{S2,t} + I_{S2,t} - u_{S2,D2,t} - E_{S2,t} - w_{S2,t} - env_{S2,t}$

(2) $S_{S1,t+1} = S_{S1,t} + I_{S1,t} - u_{S1,D2,t} - u_{S1,R,t} + u_{R,S1,t} - E_{S1,t} - w_{S1,t} - env_{S1,t}$

where $S_{k,t}$ is the storage at time t for reservoir k; $I_{k,t}$, $E_{k,t}$, $w_{k,t}$ and $env_{k,t}$ are natural inflow, evaporation, spills and compensation flow (released to meet downstream ecological flow requirements) respectively; and $u_{k,j,t}$ are controlled flows along a link between two nodes (k,j) at time t. These controlled flows include the abstractions from the two reservoirs to the demand nodes and the abstraction from the river that is pumped back into reservoir S1 (see

Figure 2 in the main manuscript). Evaporation fluxes are computed by multiplying the reservoir surface areas by the unit evaporation rate. Reservoir surface areas at each time step are calculated from storages using the available storage-elevation curves and the unit evaporation rate (assumed equal for both reservoirs) is taken from Robinson et al. (2016). Spills are calculated by imposing the hard constraint that storages at the following time-step (left hand side of Eq (1)-(2)) should never exceed the reservoir capacities, hence they are either equal to zero or to the excess volumes generated by all other terms on the right-hand side of Eq (1)-(2). Environmental compensation flows are equal to prescribed values that are constant over the year (1 Ml/day for S1 and 5 Ml/day for S2) plus occasional fish releases. Controlled fluxes are calculated via a set of operating rules, further explained below, and subject to a range of licensing and operational constraints.

The aim of the system operation is to reliably supply water while reducing pumping costs. This leads to formulating four 'daily costs', all to be minimised, shown in equations (3 - 6)

(3) Deficit for Company 1: $g_{1,t} = \max(d_{D1} - I_{R2,t} + u_{R,S1,t} - u_{S1,R,t} - q_{D1,t}, 0)$

(4) Deficit for Company 2: $g_{2,t} = \max(d_{D2,t} - u_{S1,D2,t} - u_{S2,D2,t}, 0)$

(5) Cost for Company 1: $g_{3,t} = c_{d1} * q_{D1,t} + c_{rs1} * u_{R,S1,t}$

(6) Cost for Company 2: $g_{4,t} = c_{s1d2} * u_{S1,D2,t}$

where $d_{D2,t}$ is the time-varying demand for water treatment works 2, d_{D1} is the constant demand for water treatment works 1, p_{D1} is the minimum flow required after abstraction at point D1, and c_y is the pumping costs associated with a given flow along link y. These daily costs are then translated into four objectives by taking their averages over time and over a Monte Carlo simulation ensemble, as discussed in the experimental setup section of the main manuscript.

Operation policy

Of the variables in equations (1 - 2), the controlled fluxes $u_{i,j,t}$ are the decision variables that the operators have control over, and which determine the performance of operations. To determine their values, we formulate an 'operation policy', i.e. a function that takes system state variables at the current time-step as inputs, and returns the decision variables for that time-step as outputs. For the policy, we use a Radial Basis Function Network (RBFN) in the formulation originally described in Broomhead and Lowe (1988). We visualize a single evaluation of an operation policy (as would occur every time-step) in Figure A1.1.

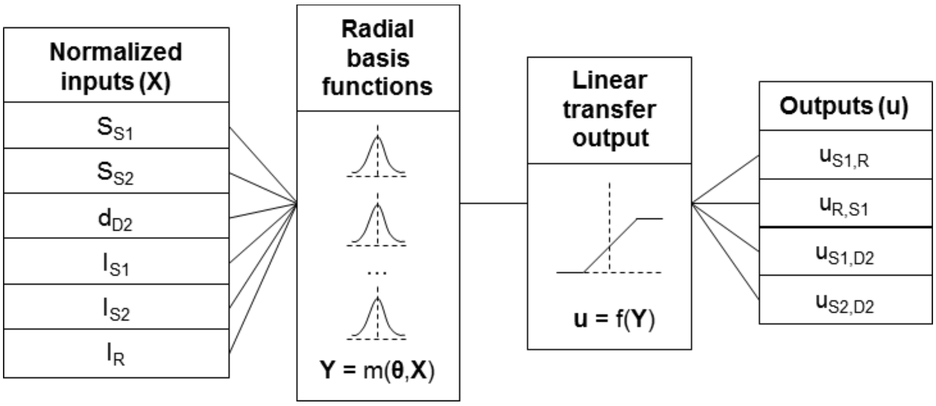


Figure A1.1. A schematic showing a single evaluation of an operation policy for a given time step. This process takes the system states at a given time-step as inputs (normalized between their maximum and minimum possible values) and returns normalized outputs that are then scaled to specify abstractions from reservoirs and rivers in the system.

The process inside the 'Radial basis functions' step is given by equation (7) below:

(7) $Y_{p,t} = a_p + \sum_{n=1}^{N} b_{n,p} q_n$, where $q_n = exp\left(-c_n \sum_{m=1}^{M} (d_{m,n} - X_{m,t})^2\right)$

where X is the vector of the network inputs (with M=6 in our case), and *a* (output biases), *b* (output weights), *c* (inverse variances) and *d* (centres) are the network parameters, which all together form the parameter vector θ used in Figures 1 and 4, N is the number hidden nodes in the network, and $Y_{p,t}$ is the network output that, after de-normalisation, becomes $u_{p,t}$. In our application we follow the rule-of-thumb described by Heaton (2008) – that the number of hidden nodes should lie between the number of inputs and the number of outputs – and use N=5.

Appendix 2: Synthetic generation of forcing inputs

The Table below lists the stochastic variables that appear in the water resource system model and the model used for their synthetic generation.

Variables	Model used for synthetic generation
Inflows: I_{R1}, I_{R2}, I_{S1}, I_{S2}	Periodic logarithmic autoregressive model
Demands: d_{D2}	Periodic autoregressive model
Unit Evaporation: ue	Periodic autoregressive model
Pump failures	Poisson duration of breaks and between breaks
Fisheries release: p_{fish}	Uniform probability of occurrence (over either September or whole year, depending on the framing) and duration

Periodic autocorrelated variables (inflows, demands or potential evaporation) are generated at each time-step following the equation:

$$X_t = \mu_t \exp(Y_t)$$

in the logarithmic case (i.e. reservoir inflows) and

$$X_t = \mu_t + Y_t$$

otherwise (demand and evaporation) where X_t represents the autocorrelated variable (I, d or ue), μ_t is the periodic component and Y_t the autocorrelated component.

The periodic component μ_t is given by the equation:

$$\mu_t = b_1 + b_2 \sin(\lambda_1 \pi f_t) + b_3 \cos(\lambda_1 \pi f_t) + \ldots + b_{N-1} \sin(\lambda_{(N-1)/2} \pi f_t) + b_N.\cos(\lambda_{(N-1)/2} \pi f_t)$$

where $f_t = (t \bmod P)/P$ with P = 365 represents the time of the year, the coefficients $\lambda_1, \lambda_2, \ldots, \lambda_{(N-1)/2}$ represent the harmonic frequencies characterising the modelled variable and the coefficients $b_1, b_2, \ldots, b_N$ are the amplitudes of those frequencies. In our application we use two harmonics for the inflows (annual and biannual) and hence set N = 5, $\lambda_1 = 2$ and $\lambda_2 = 4$;

two harmonics for the demands (annual and weekly), i.e. N = 5, λ_1=2 and λ_2=2P/7; and one harmonic (annual) for the evaporation, i.e. N = 3, λ_1=2. Once the harmonic frequencies have been set, the coefficients $b_1, b_2, \ldots, b_N$ are found using least-squares fitting of the historic data.

The autocorrelated component Y_t is given by the equation:

$$Y_t = a_0 + a_1.Y_{t-1} + a_2.Y_{t-2} + \ldots + a_L.Y_{t-L} + \varepsilon_t,$$

where a_0 is the expected value of Y_t and the coefficients $a_1, a_2, \ldots, a_L$ represent the lagged correlations and ε represents the 'innovation' (a normal random variable). The number of lag terms (L) takes the value 1 or 2 depending on the framing. For given L, the coefficients $a_1, a_2, \ldots, a_L$ are determined by a least-squares fitting of historic data using. Since ε is correlated across variables, we transform uncorrelated random normal numbers by the Cholesky decomposition of the variance-covariance matrix of ε (found from the historic data) to create correlated normal variables, as described in (Gentle, 2009).

Pump failures are generated using two Poisson distributions, one describing the duration between breaks and one describing the duration of each break. The expected duration between breaks is 300 days for $u_{S1,D2}$ and $u_{R,S1}$, and 800 days for $u_{S2,D2}$. The expected duration of a break is 3 days for $u_{S1,D2}$ and $u_{R,S1}$, and 5 days for $u_{S2,D2}$. We note that, although $u_{S2,D2}$ is not a pumped flow, pump failures represent any failure to use the pipe, thus we suggest that a failure to use this link is possible (albeit less likely). We assume that, as a direct dam release, $u_{S1,R}$ is always possible.

Fisheries releases are assumed to happen at most once a year. The release event must occur in a certain time window that, depending on the framing, either covers the all year (except Spring) or spans over September only. Each day in this window has an equal chance of being the beginning of the fisheries release. Once the first release day has been randomly extracted, the release duration may last from 3 to 14 days (this duration is also randomly selected). The specified volume (900 Ml) is then uniformly released over the duration of the release.

Appendix 3: Supporting results

Pre-optimization

As established in Section 3.3, an appropriate sample size needs to be selected to ensure that the ‘approximation error’ in each objective is smaller than the sensitivity of the decision maker to variations in that objective. In Figure 6 we demonstrate how to determine the sample size for one objective in one framing. In Figure A3.1 we show these plots for all objectives and all framings. To create this figure, we evaluated 132 random operating policies over increasingly large sample sizes for simulation and recorded their objective values. We then recorded the absolute difference between an objective value at a given sample size and the objective value at the largest sample size ($1.2*10^7$ days), this difference is the ‘approximation error’. The lines in this plot indicate the 99th percentile (across the 132 random policies) of the approximation error for a given framing. These plots enable us to choose a sample size that limits the approximation error to a specified value (indicated by the dotted lines). There are 8 lines because we use 8 framings. From these plots we see that a sample size of $1.2*10^7$ days is acceptable because the 99th percentile approximation error for all framings in all objectives is less than the decision maker’s sensitivity to variations in that objective.

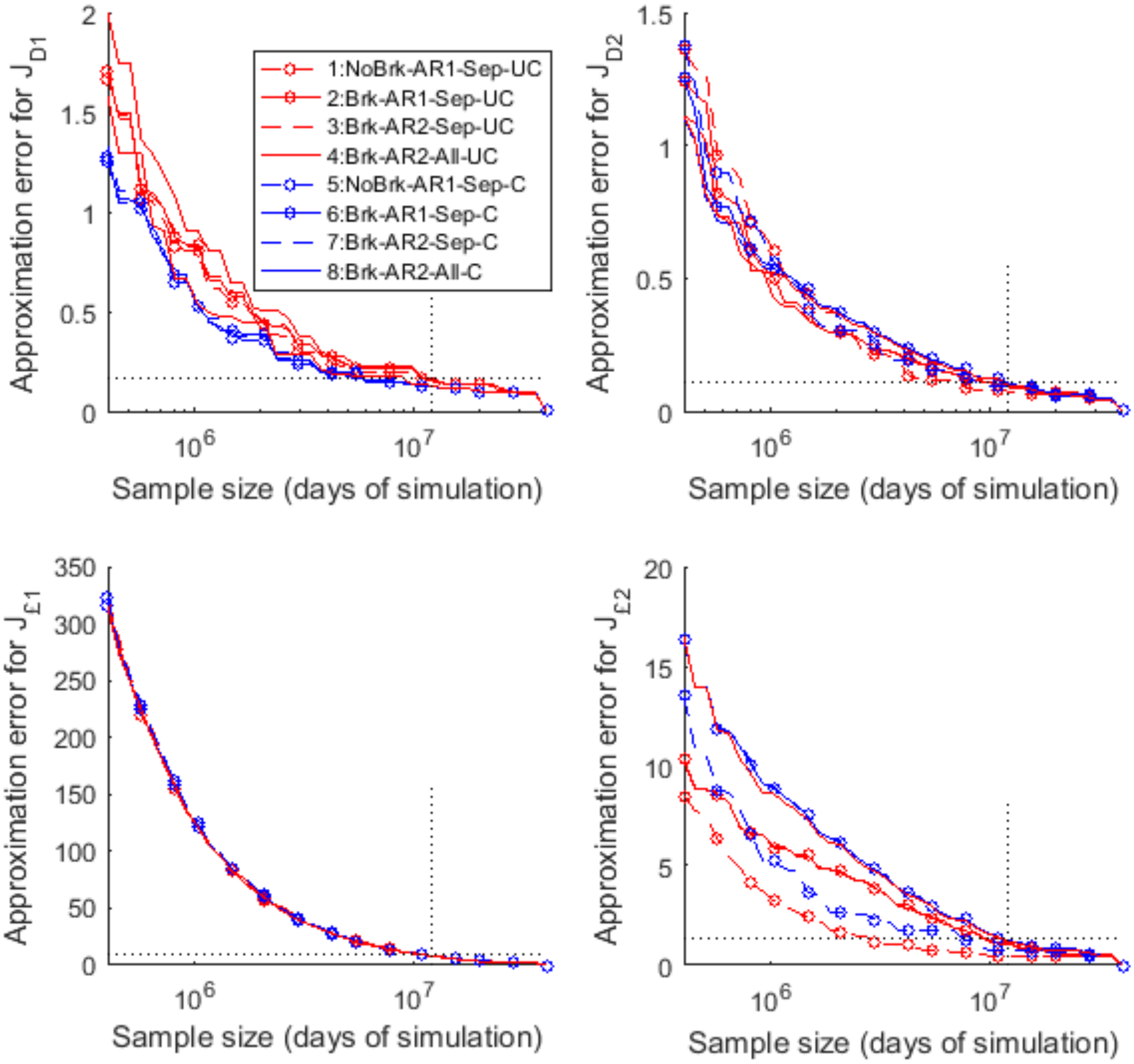


Figure A3.1. Each coloured line marks the trajectory of the 99th percentile of error for a given objective, in a given framing. The black dotted lines mark the selected sample size, K*T, ($1.2*10^7$) and the approximation error which we also define as the epsilon value.

To determine how many function evaluations are required for the optimization process, it is important to find the point at which including more function evaluations will provide a negligible benefit. To do this we performed optimization (in framing 8) 25 times, recording the objective values of the population of operating policies every 100 iterations of the optimization. These objective values are then normalized between 0 and 1. The hypervolume that is dominated by these objective values is plotted in the plot in Figure A3.2 below. This plot suggests that 10^5 function evaluations in the optimization process should be more than sufficient.

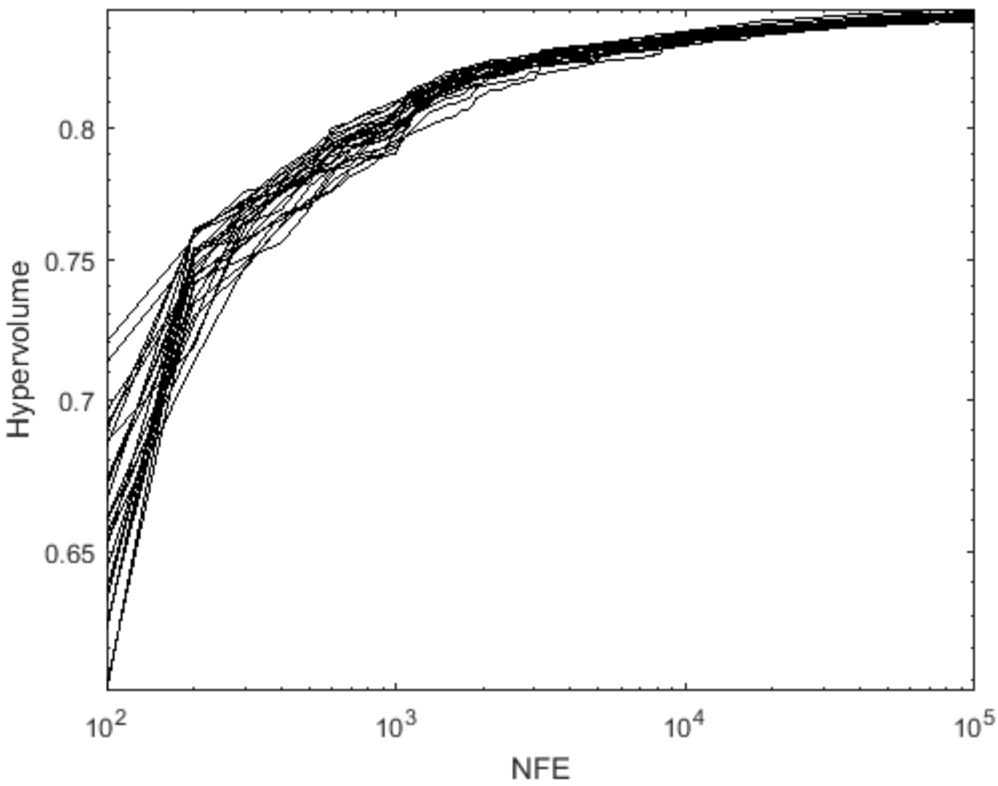


Figure A3.2. The hypervolume trajectory of 25 optimization runs up to 10^5 Number of Function Evaluations (NFE), each with a different seed for the Borg MOEA and evaluated on forcing with different seeds. See (Salazar et al., 2016) for a more detailed discussion.

Results

Figure A3.3 presents the full results of the re-evaluation step of our proposed workflow. To create Figure A3.3, we take sets of Pareto solutions (where each solution is an operating policy of our reservoir system) that have been optimized to a given framing (as indicated by the legend) and re-evaluate them in other framings (as indicated by the row number). We then plot the Cumulative Distribution Function (CDF) of the differences in objective values (each column shows a different objective) between optimization and re-evaluation. This enables us to visualise how estimates of objective values change under different realizations of uncertainty, i.e. under framings that are different from those used in optimization. CDFs that lie to the right of the X=0 line perform better in the framing used for re-evaluation than in the framing used for optimization, and vice versa if they lie to the left of the X=0 line.

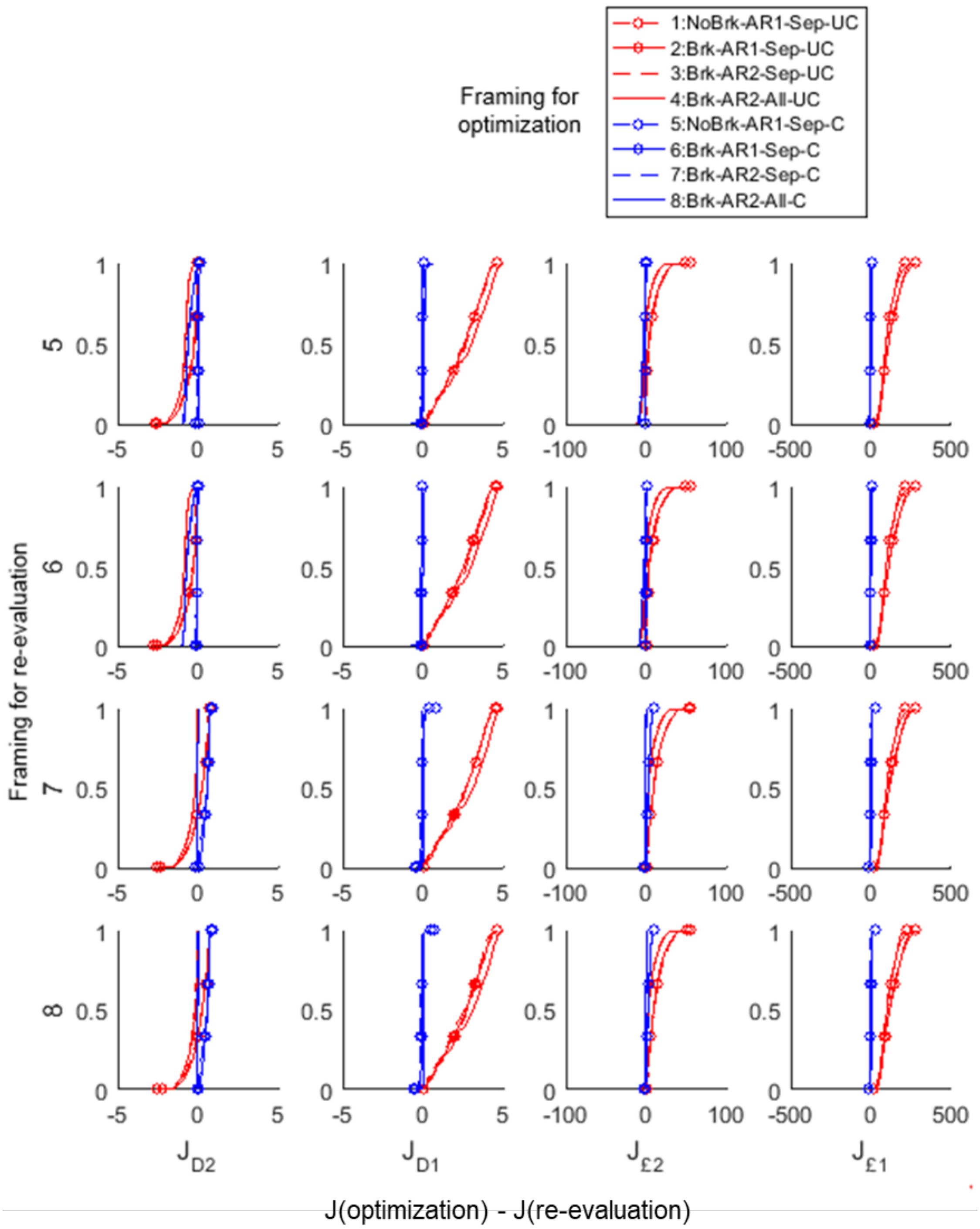


Figure A3.3. CDFs of the difference between re-evaluation and calibration of an objective (positive indicates improvement on re-evaluation and negative indicates deterioration). Each line indicates the framing number as given in the legend, cooperative framings are shown in blue and non-cooperative in red.

Figure A3.4 shows the same as Figure A3.3 but using a smaller (30 year, 10^4 days) sample size. This figure should be contrasted with Figure A3.3 to show that there are no consistent patterns when a small sample size is used – the differences between framings appear to be random (as opposed to the consistency seen in Figure A3.3).

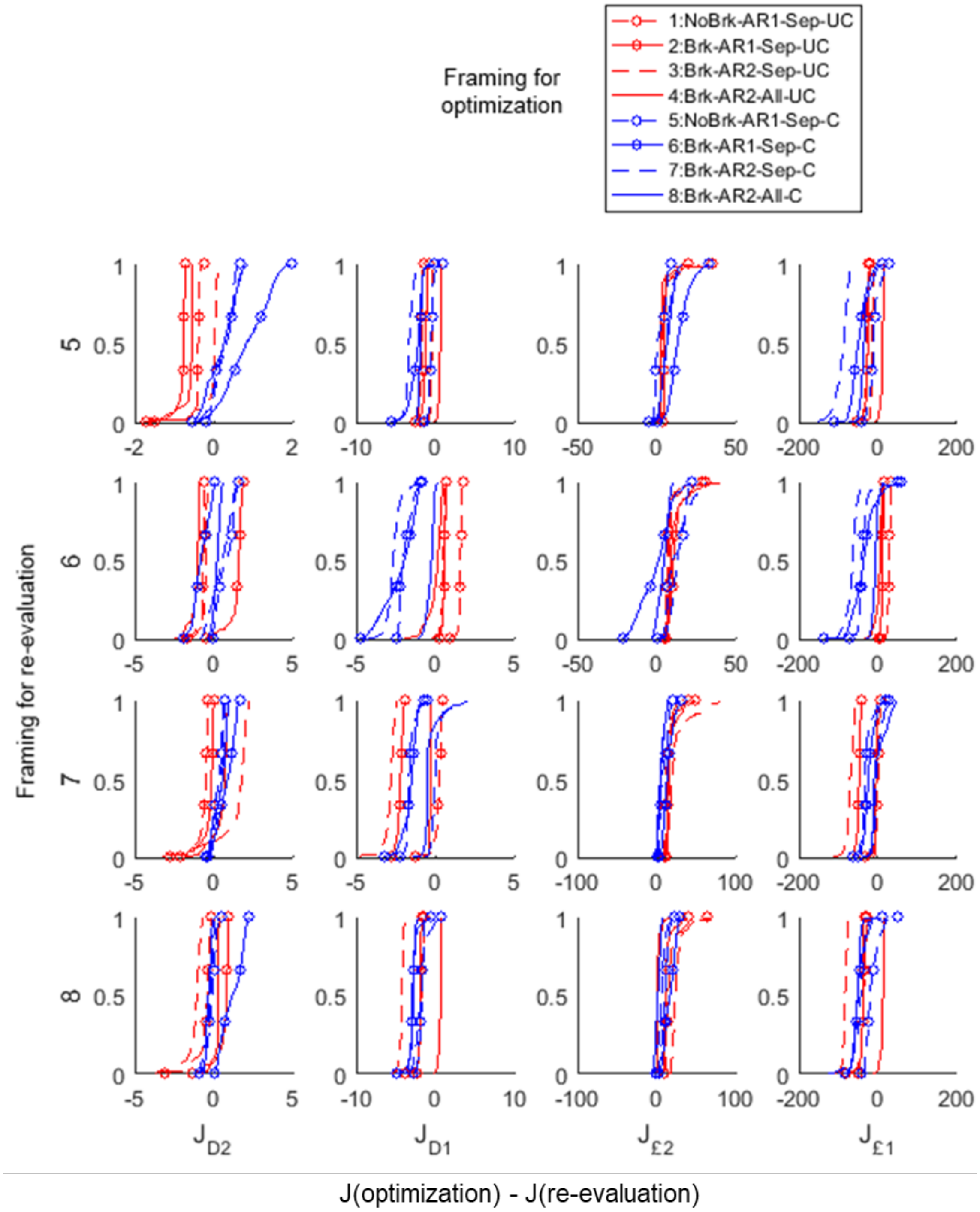


Figure A3.4. Same as in Figure A3.3 but using a much shorter (30 year) simulation length in both optimization and re-evaluation.

References

Anghileri, D., A. Castelletti, F. Pianosi, R. Soncini-Sessa, and E. Weber (2013), Optimizing Watershed Management by Coordinated Operation of Storing Facilities, *Journal of Water Resources Planning and Management*, *139*(5), 492-500.

Ayala, A. I., A. Cortés, W. E. Fleenor, and F. J. Rueda (2014), Seasonal scale modeling of river inflows in stratified reservoirs: Structural vs. parametric uncertainty in inflow mixing, *Environmental Modelling & Software*, *60*, 84-98.

Beck, M. B. (1987), Water quality modeling: A review of the analysis of uncertainty, *Water Resources Research*, *23*(8), 1393-1442.

Beven, K. J., and R. E. Alcock (2012), Modelling everything everywhere: a new approach to decision-making for water management under uncertainty, *Freshwater Biology*, *57*, 124-132.

Beven, K. J., et al. (2017), Epistemic uncertainties and natural hazard risk assessment. 1. A review of different natural hazard areas, *Natural Hazards and Earth System Sciences Discussions*, 1-53.

Broomhead, D. S., and D. Lowe (1988), Radial basis functions, multi-variable functional interpolation and adaptive networks, Royal Signals and Radar Establishment Malvern (United Kingdom).

Brown, C., J. R. Lund, X. Cai, P. M. Reed, E. A. Zagona, A. Ostfeld, J. Hall, G. W. Characklis, W. Yu, and L. Brekke (2015), The future of water resources systems analysis: Toward a scientific framework for sustainable water management, *Water Resources Research*, *51*(8), 6110-6124.

Castelletti, A., F. Pianosi, and R. Soncini-Sessa (2012), Stochastic and Robust Control of Water Resource Systems: Concepts, Methods and Applications, 383-401.

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resources Research*, *44*(12).

Cohon, J. L., and D. H. Marks (1975), A review and evaluation of multiobjective programing techniques, *Water Resources Research*, *11*(2), 208-220.

Donkor, E. A., T. A. Mazzuchi, R. Soyer, and J. Alan Roberson (2014), Urban Water Demand Forecasting: Review of Methods and Models, *Journal of Water Resources Planning and Management*, *140*(2), 146-159.

Dooge, J. (1973), *Linear theory of hydrologic systems*, Agricultural Research Service, US Department of Agriculture.

EA, E. A. (2017), Large Raised Reservoirs (AfA134), edited by E. A. EA, data.gov.uk, Online.

Fiering, M. B., and B. Bund (1971), *Synthetic streamflows*, American Geophysical Union.

Fitzpatrick, J. M., and J. J. Grefenstette (1988), Genetic Algorithms in Noisy Environments, *Machine Learning*, *3*(2/3), 101-120.

Gentle, J. E. (2009), *Computational statistics*, Springer Science & Business Media.

Giuliani, M., and A. Castelletti (2013), Assessing the value of cooperation and information exchange in large water resources systems by agent-based optimization, *Water Resources Research*, *49*(7), 3912-3926.

Giuliani, M., A. Castelletti, F. Amigoni, and X. Cai (2015a), Multiagent Systems and Distributed Constraint Reasoning for Regulatory Mechanism Design in Water Management, *Journal of Water Resources Planning and Management*, *141*(4).

Giuliani, M., A. Castelletti, F. Pianosi, E. Mason, and P. M. Reed (2015b), Curses, Tradeoffs, and Scalable Management: Advancing Evolutionary Multiobjective Direct Policy Search to Improve Water Reservoir Operations, *Journal of Water Resources Planning and Management*, 04015050.

Guariso, G., S. Rinaldi, and R. Soncini-Sessa (1986), The Management of Lake Como: A Multiobjective Analysis, *Water Resources Research*, *22*(2), 109-120.

Hadka, D., and P. Reed (2013), Borg: an auto-adaptive many-objective evolutionary computing framework, *Evol Comput*, *21*(2), 231-259.

Haimes, Y. Y., and W. A. Hall (1977), Sensitivity, responsivity, stability and irreversibility as multiple objectives in civil systems, *Advances in Water Resources*, *1*(2), 71-81.

Hashimoto, T., J. R. Stedinger, and D. P. Loucks (1982), Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation, *Water Resources Research*, *18*(1), 14-20.

Heaton, J. (2008), *Introduction to neural networks with Java*, Heaton Research, Inc.

Herman, J. D., H. B. Zeff, P. M. Reed, and G. W. Characklis (2014), Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty, *Water Resources Research*, *50*(10), 7692-7713.

Herman, J. D., H. B. Zeff, J. R. Lamontagne, P. M. Reed, and G. W. Characklis (2016), Synthetic Drought Scenario Generation to Support Bottom-Up Water Supply Vulnerability Assessments, *Journal of Water Resources Planning and Management*, *142*(11).

Hiew, K. L., J. W. Labadie, and J. F. Scott (1989), Optimal operational analysis of the Colorado-Big Thompson project, paper presented at Computerized decision support systems for water managers, ASCE.

Hirsch, R. M. (1979), Synthetic hydrology and water supply reliability, *Water Resources Research*, *15*(6), 1603-1615.

Höllermann, B., and M. Evers (2017), Perception and handling of uncertainties in water management—A study of practitioners' and scientists' perspectives on uncertainty in their daily decision-making, *Environmental Science & Policy*, *71*, 9-18.

Hutton, C. J., and Z. Kapelan (2015), A probabilistic methodology for quantifying, diagnosing and reducing model structural and predictive errors in short term water demand forecasting, *Environmental Modelling & Software*, *66*, 87-97.

ICoLD (2003), World Register of Dams, edited by I. C. o. L. Dams, Paris.

Kasprzyk, J. R., P. M. Reed, G. W. Characklis, and B. R. Kirsch (2012), Many-objective de Novo water supply portfolio planning under deep uncertainty, *Environmental Modelling & Software*, *34*, 87-104.

Kasprzyk, J. R., S. Nataraj, P. M. Reed, and R. J. Lempert (2013), Many objective robust decision making for complex environmental systems undergoing change, *Environmental Modelling & Software*, *42*, 55-71.

Knowles, J., and D. Corne (2002), On metrics for comparing nondominated sets, paper presented at Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on, 12-17 May 2002.

Koutsoyiannis, D., and A. Economou (2003), Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems, *Water Resources Research*, *39*(6).

Labadie, J. W. (2004), Optimal Operation of Multireservoir Systems: State-of-the-Art Review, *Journal of Water Resources Planning and Management*, *130*(2), 93-111.

Langsdale, S., A. Beall, E. Bourget, E. Hagen, S. Kudlas, R. Palmer, D. Tate, and W. Werick (2013), Collaborative Modeling for Decision Support in Water Resources: Principles and Best Practices, *JAWRA Journal of the American Water Resources Association*, *49*(3), 629-638.

Lempert, R., S. Popper, and S. C. Bankes (2003), *Shaping the next one hundred years: new methods for quantitative, long-term policy analysis*, Rand Corporation.

Lempert, R., D. G. Groves, S. W. Popper, and S. C. Bankes (2006), A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios, *Management Science*, *52*(4), 514-528.

Maass, A., M. M. Hufschmidt, R. Dorfman, H. A. Thomas Jr, S. A. Marglin, and G. M. Fair (1962), *Design of water-resource systems*, Harvard University Press Cambridge, Massachusetts.

Mahmoud, M., et al. (2009), A formal framework for scenario development in support of environmental decision-making, *Environmental Modelling & Software*, *24*(7), 798-808.

Marques, G. F., and A. Tilmant (2013), The economic value of coordination in large-scale multireservoir systems: The Parana River case, *Water Resources Research*, *49*(11), 7546-7557.

Matalas, N. C. (1967), Mathematical assessment of synthetic hydrology, *Water Resources Research*, *3*(4), 937-945.

Miller, B. L., and D. E. Goldberg (1996), Optimal sampling for genetic algorithms, *Urbana*, *51*, 61801.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008), Stationarity Is Dead: Whither Water Management?, *Science*, *319*(5863), 573-574.

Montanari, A., and D. Koutsoyiannis (2014), Modeling and mitigating natural hazards: Stationarity is immortal!, *Water Resources Research*, *50*(12), 9748-9756.

Mortazavi, M., G. Kuczera, and L. Cui (2012), Multiobjective optimization of urban water resources: Moving toward more practical solutions, *Water Resources Research*, *48*(3).

Moss, R., et al. (2016), Understanding Dynamics and Resilience in Complex Interdependent Systems, in *U.S. Global Change Research Program Interagency Group on Integrative Modeling*, edited, Washington, D.C.

Nardini, A., C. Piccardi, and R. Soncini-Sessa (1992), On the integration of risk aversion and average-performance optimization in reservoir control, *Water Resources Research*, *28*(2), 487-497.

Oliveira, R., and D. P. Loucks (1997), Operating rules for multireservoir systems, *Water Resources Research*, *33*(4), 839-852.

Pianosi, F., X. Q. Thi, and R. Soncini-Sessa (2011), Artificial Neural Networks and Multi Objective Genetic Algorithms for Water Resources Management: An Application to the Hoabinh Reservoir in Vietnam, 10579-10584.

Quinn, J. D., P. M. Reed, M. Giuliani, and A. Castelletti (2017), Rival framings: A framework for discovering how problem formulation uncertainties shape risk management trade-offs in water resources systems, *Water Resources Research*, *53*(8), 7208-7233.

Rajagopalan, B., J. D. Salas, and U. Lall (2010), STOCHASTIC METHODS FOR MODELING PRECIPITATION AND STREAMFLOW, in *Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting*, edited, pp. 17-52, WORLD SCIENTIFIC.

Rani, D., and M. M. Moreira (2009), Simulation–Optimization Modeling: A Survey and Potential Application in Reservoir Systems Operation, *Water Resources Management*, *24*(6), 1107-1138.

Reed, P. M., and J. B. Kollat (2013), Visual analytics clarify the scalability and effectiveness of massively parallel many-objective optimization: A groundwater monitoring design example, *Advances in Water Resources*, *56*, 1-13.

Robinson, E. L., E. Blyth, D. B. Clark, E. Comyn-Platt, J. Finch, and A. C. Rudd (2016), Climate hydrology and ecology research support system potential evapotranspiration dataset for Great Britain (1961-2015) [CHESS-PE], edited, NERC Environmental Information Data Centre.

Rogers, P. P., and M. B. Fiering (1986), Use of systems analysis in water management, *Water Resources Research*, *22*(9S), 146S-158S.

Rosenstein, M. T., and A. G. Barto (2001), Robot weightlifting by direct policy search, paper presented at International Joint Conference on Artificial Intelligence, Citeseer.

Salas, J. D., and J. T. B. Obeysekera (1982), ARMA Model identification of hydrologic time series, *Water Resources Research*, *18*(4), 1011-1021.

Salas, J. D., C. Fu, A. Cancelliere, D. Dustin, D. Bode, A. Pineda, and E. Vincent (2005), Characterizing the Severity and Risk of Drought in the Poudre River, Colorado, *Journal of Water Resources Planning and Management*, *131*(5), 383-393.

Salazar, J. Z., P. M. Reed, J. D. Herman, M. Giuliani, and A. Castelletti (2016), A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control, *Advances in Water Resources*, *92*, 172-185.

Salazar, J. Z., P. M. Reed, J. D. Quinn, M. Giuliani, and A. Castelletti (2017), Balancing exploration, uncertainty and computational demands in many objective reservoir optimization, *Advances in Water Resources*, *109*, 196-210.

Simonovic, S. P. (1992), Reservoir Systems Analysis: Closing Gap between Theory and Practice, *Journal of Water Resources Planning and Management*, *118*(3), 262-280.

Smalley, J. B., B. S. Minsker, and D. E. Goldberg (2000), Risk-based in situ bioremediation design using a noisy genetic algorithm, *Water Resources Research*, *36*(10), 3043-3052.

Stedinger, J. R., B. F. Sule, and D. P. Loucks (1984), Stochastic dynamic programming models for reservoir operation optimization, *Water Resources Research*, *20*(11), 1499-1505.

Tejada-Guibert, J. A., S. A. Johnson, and J. R. Stedinger (1995), The Value of Hydrologic Information in Stochastic Dynamic Programming Models of a Multireservoir System, *Water Resources Research*, *31*(10), 2571-2579.

Tilmant, A., and W. Kinzelbach (2012), The cost of noncooperation in international river basins, *Water Resources Research*, *48*(1).

Trindade, B. C., P. M. Reed, J. D. Herman, H. B. Zeff, and G. W. Characklis (2017), Reducing regional drought vulnerabilities and multi-city robustness conflicts using many-objective optimization under deep uncertainty, *Advances in Water Resources*, *104*, 195-209.

Vogel, R. M. (2017), Stochastic watershed models for hydrologic risk management, *Water Security*, *1*, 28-35.

Walker, W. E., P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen, and M. P. Krayer von Krauss (2003), Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support, *Integrated Assessment*, *4*(1), 5-17.

Watson, A. A., and J. R. Kasprzyk (2017), Incorporating deeply uncertain factors into the many objective search process, *Environmental Modelling & Software*, *89*, 159-171.

Weiss, H. K. (1956), Estimation of Reliability Growth in a Complex System with a Poisson-Type Failure, *Operations Research*, *4*(5), 532-545.

Wu, X., C. Cheng, Y. Zeng, and J. R. Lund (2016), Centralized versus Distributed Cooperative Operating Rules for Multiple Cascaded Hydropower Reservoirs, *Journal of Water Resources Planning and Management*, *142*(11).

Yakowitz, S. (1982), Dynamic programming applications in water resources, *Water Resources Research*, *18*(4), 673-696.

Yeh, W. W. G. (1985), Reservoir Management and Operations Models: A State-of-the-Art Review, *Water Resources Research*, *21*(12), 1797-1818.

Yun, R., V. P. Singh, and Z. Dong (2010), Long-Term Stochastic Reservoir Operation Using a Noisy Genetic Algorithm, *Water Resources Management*, *24*(12), 3159-3172.